FOOD SECURITY IN PRACTICE

Methods for Rural Development Projects

Edited by John Hoddinott

International Food Policy Research Institute
Washington, DC

ISBN 0-89629-713-6
Library of Congress Cataloging-in-Publication data available.

Contents

Tables, Figures, and Boxes

Tables

Figures

Boxes

Foreword

The International Food Policy Research Institute (IFPRI), like many development practitioners, often finds that insufficient information constrains its research efforts with partner organizations in developing countries. Solid data are often lacking on the nature of poverty, food insecurity, and malnutrition; the location of food insecure areas; and the causal links between potential interventions and outcomes of interest. This absence of information adversely affects the design, implementation, monitoring, and evaluation of interventions, including those designed to ameliorate food insecurity and malnutrition.

This book, based on IFPRI's field experience and interaction with a variety of partner organizations, aims to assist development practitioners in overcoming these constraints. The principal audience is an operational one—multilateral or bilateral aid agencies, nongovernmental organizations (NGOs), developing-country governments, and other development practitioners actively engaged in food security and nutrition issues. The book provides a framework for thinking about what projects would be most appropriate in a given situation and indicates what types of information are needed in order to maximize project impact. It can also assist by making development practitioners more fully conversant with food security and nutrition concepts.

The authors have sought to make this book "fieldwork friendly," so it is not intended as an exhaustive survey of all issues or methods with respect to food security and nutrition. Rather, the material presented here is designed to provide a suite of useful methods relevant at different points in the project cycle. Although each chapter stands alone, I encourage readers to begin with the introduction, which provides an overview of the key issues.

IFPRI's mission is to search for policies to feed the world and protect the environment. I hope that this guide will assist others who share our goal by facilitating the targeting and design of interventions for maximum effect on food insecurity and nutrition, and by facilitating the development of better methods for monitoring and evaluation.

Per Pinstrup-Andersen
Director General, IFPRI

Acknowledgments

Much of the material presented here was originally produced for the International Fund for Agricultural Development (IFAD) in the form of a series of technical guides for operationalizing household food security. Specifically, funding for data collection and analysis was supported by IFAD's Technical Assistance Grant 301–IFPRI. In addition, this work has been aided by comments and advice we received from a number of IFAD staff: Mona Fikry, Shantanu Mathur, Annina Lubbock, Enrique Murguia Oropeza, Sana Jatta, Gary Howe, and David Kingsbury. The authors gratefully acknowledge this funding.

Some of the material also draws on work undertaken for The United States Agency for International Development (USAID) under grants 688-C-00-98-00151-00 and 698-0478-G-00-5-272-00. In particular, we would like to acknowledge, with thanks, the support of Roger Bloom and Kevin Sturr.

We would also like to record our appreciation to the research staff who worked with us in the field. It is a pleasure to commend the outstanding work of Juan Manuel Medina Banegas (Honduras), Luc Christiaensen (Mali), Sidi Guindo (Mali), Abdourhamane Maiga (Mali), and Charles Masangano (Malawi). In Washington, we were ably assisted by Lynette Aspillera, Ginette Mignot, Jay Willis, and Yisehac Yohannes. We would also like to acknowledge the guidance and enthusiasm of Lawrence Haddad and the valuable contribution of Detlev Puetz.

Most important, the data used in this volume came from hundreds of men and women in Honduras, Malawi, and Mali who, with good humor, patiently answered many questions on their own food security and nutrition. We recognize that the hours they spent with us represented a genuine opportunity cost; we hope that in some way they will benefit from this work.

The ideas and opinions presented in this volume are the sole responsibility of the authors.

1. Introduction

John Hoddinott

This book is principally aimed at individuals in multilateral or bilateral aid agencies, nongovernmental organizations (NGOs), developing-country governments, and other development practitioners who are actively engaged in food security or nutrition issues. These practitioners often are knowledgeable about general development issues and have substantial managerial prowess, but lack materials that could provide a bridge between the academic literature on these issues and the operational concerns associated with designing, implementing, monitoring, and evaluating projects.

The purpose of this book is to help bridge this gulf between theory and practice. To begin with, project staff often face information and resource constraints. That is, information is often lacking on the nature of the food security and nutrition problems facing a country, or region within a country; the location of food-insecure areas; and the causal links between potential interventions and food security outcomes. Further, there is neither the time nor money to launch detailed, lengthy, quantitative household surveys. Even if such surveys could be launched, it is simply not feasible to apply sophisticated statistical analyses to these data.

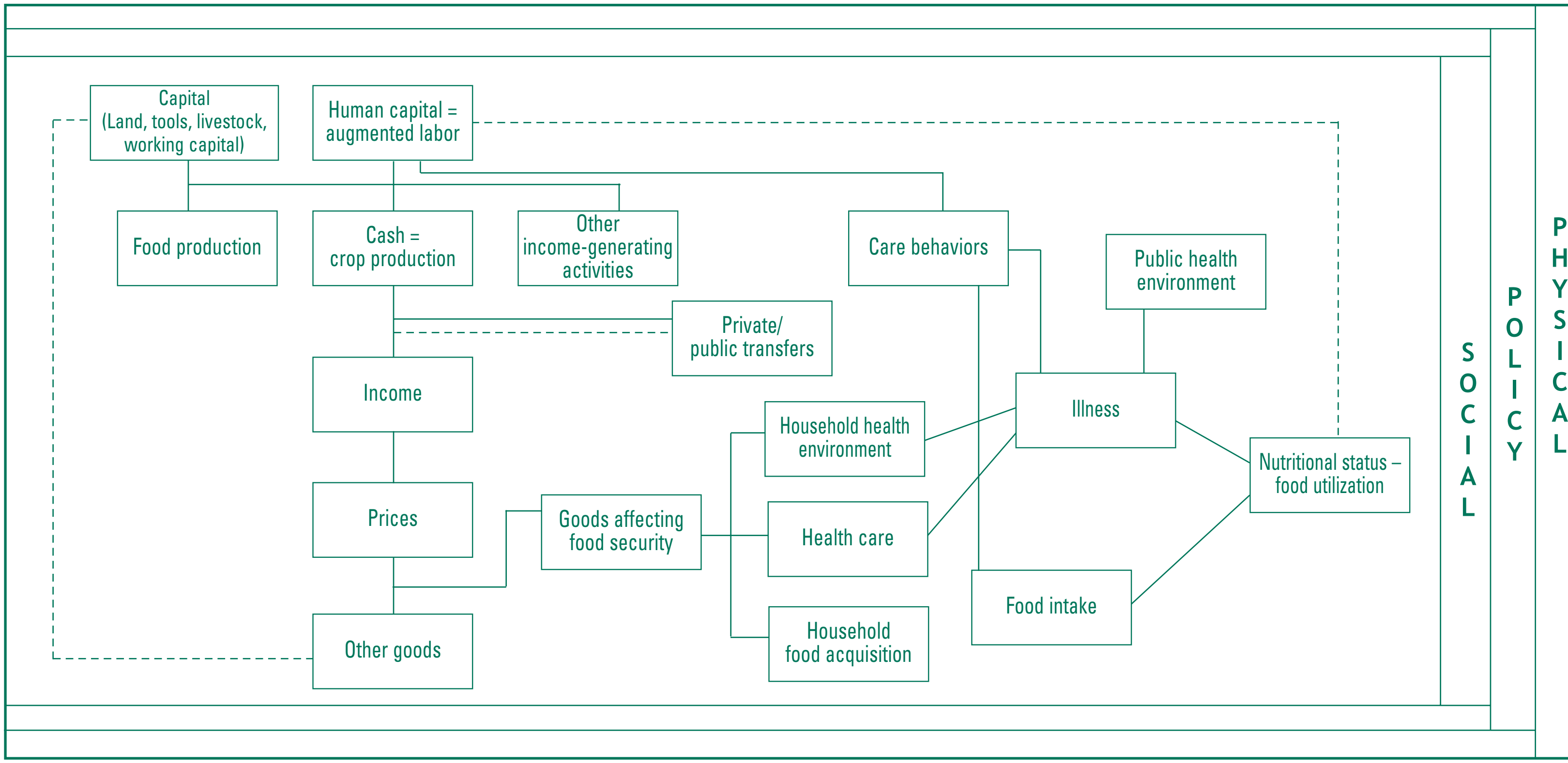
The material presented here recognizes these constraints. One of our objectives is to outline a number of relatively quick methods for obtaining information on food security and nutrition. A second is to keep the statistical requirements associated with using these data to a minimum. To apply the material and methods discussed in this report, all that is needed is access to a spreadsheet program such as Excel or Quattro and a rudimentary understanding of a few statistical techniques, such as computing means, testing hypotheses, and estimating simple linear regression models. Third, we show how project designers can use this information to understand the nature of food security and nutrition problems, target interventions more effectively, and develop simple but effective tools for monitoring and evaluation. Fourth, we try to avoid using jargon or technical language; where we do, we define these terms in a way we hope is accessible. Finally, alongside the presentation of these methods, we present examples to make the material more accessible.

This introduction attempts to do two things. First, it provides a brief introduction to the concept of food security. (An introduction to nutrition concepts and issues is found in Chapter 2.) It outlines the links between a variety of development projects and their impact on food security and nutrition. By doing so, it provides a framework for thinking about what projects would be most appropriate in a given situation and indicates what types of information are needed in order to maximize impacts on food security. Second, it introduces the material in this book, showing how it can assist staff in easing the information constraints they often face. By doing so, it should be possible to improve the targeting of interventions, understand their likely effects, and develop improved monitoring and evaluation methods.

The Links Between Development Interventions, Household Food Security, and Nutrition

Having established the relevant dimensions of food security, the next step is to outline a framework that links the concepts of food security

Figure 1.1 The determinants of household food security



Source: Developed by author from Maxwell and Frankenberger (1992, 25).

to typical development interventions. This is shown in Figures 1.1 and 1.2. We begin with Figure 1.1. As it is a little complicated, it is helpful to consider it in several steps.

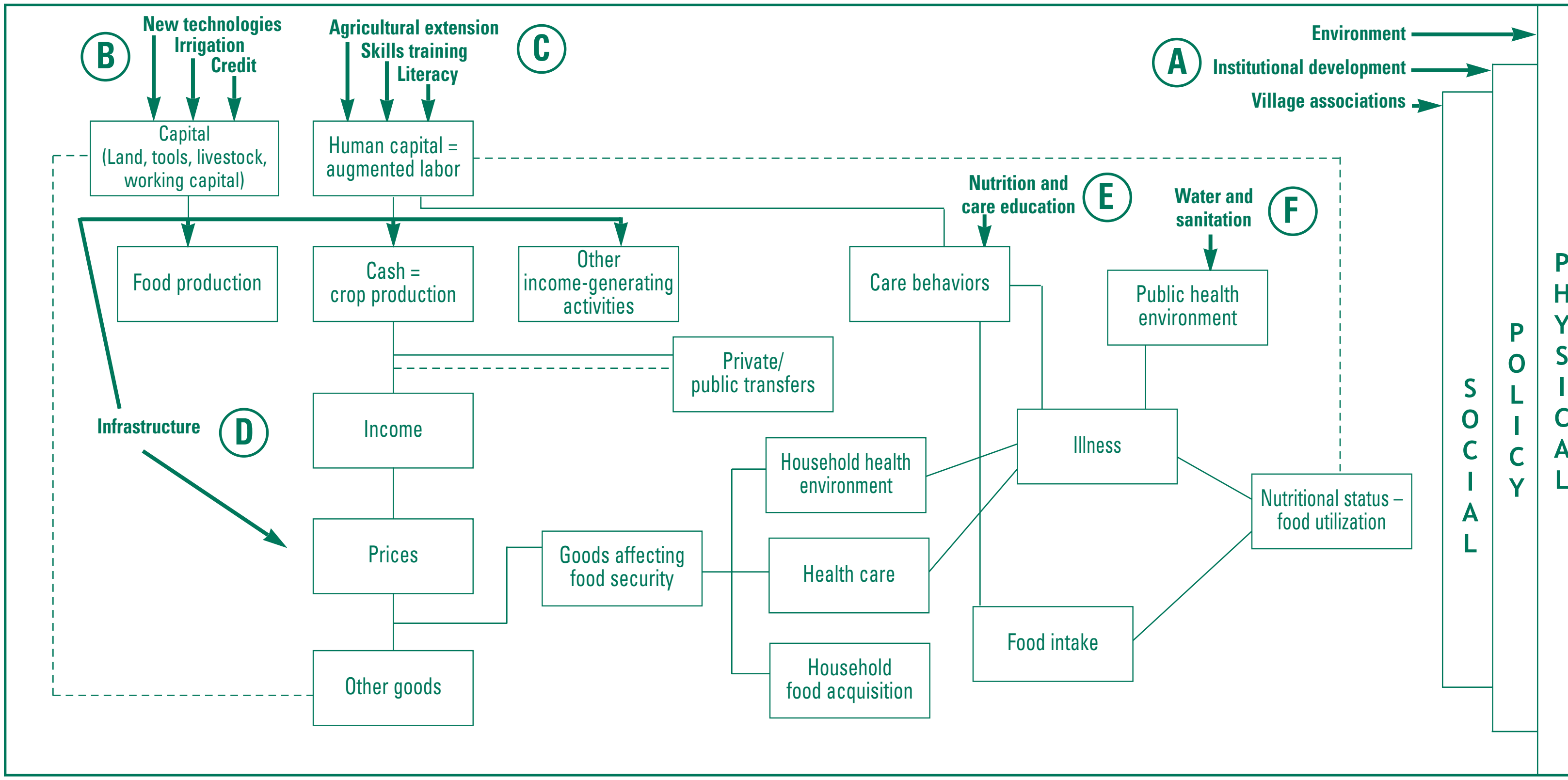
1. The diagram is "framed" by the physical, policy, and social environment. The purpose of this framing is to remind the analyst that household food security issues cannot be seen in isolation from broader factors. Examples of these "environmental" issues are as follows:

- The physical environment plays a large role in determining the type of activities that can be undertaken by rural households.
- Government policies toward the agricultural sector will have a strong effect on the design and implementation of household food security interventions. For example, a pricing policy that is hostile toward agriculture will discourage production. Interventions that ignore this fact are unlikely to succeed. The presence of social conflict, expressed in terms of mistrust of other social groups or even outright violence, is also an important factor in the design and implementation of

interventions. In such circumstances, maximizing beneficiary participation becomes especially problematic. For example, wealthier groups may take control of projects for their own benefit, to the exclusion of poorer members. Alternatively, social conflict may encourage groups excluded from an intervention to take active steps to subvert it. A certain degree of social cohesion is necessary if group activities, such as group-based microcredit schemes or collective work on an infrastructure, are to succeed.

2. The resources, or endowments, of households can be divided into two broad categories: labor and capital. Labor refers to the availability of labor for production. It incorporates both a physical dimension—how many people are available to work—as well as a "knowledge" or human capital dimension. For an agricultural household, this knowledge includes formal schooling and formal training in agricultural production. It also includes informal knowledge obtained via trial and error, past farming experiences, discussions with friends and relatives, observations made about practices on neighbors' farms, and so on. Capital refers to those resources—such as land, tools for agricultural and nonagricultural production, livestock, and financial resources—that, when combined with labor, produce income.
3. Households allocate these endowments across different activities such as food production, cash crop production, and nonagricultural income-generating activities (such as wage labor, handicrafts, food processing, services, and so on) in response to the returns each activity generates. In addition, households may receive transfer income from other households, from some public body such as the state, or from an NGO. Together, these four sources determine household income.
4. Households face a set of prices that determine the level of consumption that can be supported by this level of income.
5. Consumption is divided between those goods that affect household and individual food security and all other goods.
6. Goods that affect food security include food consumption at the household level (referred to as food access in much of the food security literature); goods directly related to health care, such as medicines; and goods that affect the health environment, such as shelter, sanitation, and water. These three goods, together with knowledge and practice of good nutritional and health practices (called "care behaviors") and the public health environment (for example, the availability of publicly provided potable water), affect illness and individual food intake, which in turn generates nutritional status or food utilization. Note that this part of the diagram is exactly the same as a diagram describing the causes of malnutrition found in Maxwell and Frankenberger (1992, 25). Stars are placed beside the household food acquisition, food intake, and food utilization boxes to emphasize that these are food security and nutrition outcomes.
7. Finally, note that food security is not static over time. There are second-round, or feedback effects, denoted by the dashed lines in Figure 1.1. Suppose a donor funds a project that improves the provision of agricultural extension. This can be thought of as a project that increases the human capital of the household. In turn, this raises income. Some of this income might be used to acquire additional capital stock, such as agricultural implements. In turn, this raises household income in subsequent years. Allocations of food, expenditures on education, and health will affect the level and distribution of human capital within the

Figure 1.2 The impact of development interventions on household food security



Source: Developed by author from Maxwell and Frankenberger (1992, 25).

household. These investments will also affect the household's ability to generate income in subsequent years. In other words, a well-designed intervention has the potential to set in train a virtuous circle of development, whereby increased income generates greater wealth, which in turn generates higher levels of income, consumption, food security, and nutrition. But it is also worth noting that not all these feedback effects are benign. Increased income generation may induce an offsetting reduction in private transfers received from other households, a phenomenon known as "crowding out."

It is now possible to uncover the links between development projects and household or individual food security. In Figure 1.2, these interventions (written in bold) are superimposed on Figure 1.1. They are placed within the diagram at the point where their direct impact is observed.

A. A series of interventions are designed to improve the broader environments that affect household food security. Examples of these include: in the environment area, field operations such as soil, water, and forest management; in the policy area, providing an appropriate institutional environment for private agriculture;

and, in the social area, strengthening small farmers' associations.

B. There are interventions that increase the level of and returns to capital. Examples include the rehabilitation of irrigation facilities, the provision of credit, and the development of new technologies.

C. There are interventions that increase the stock of knowledge or human capital. Examples include literacy training or extension services that provide new technical skills in the nonagricultural sector.

D. There are interventions that improve rural infrastructures, notably roads. Reducing transport costs improves household food security in two ways: by increasing the returns from production activities and by reducing the costs of obtaining food and other goods for consumption.

E. There are interventions to improve knowledge of good health care and nutrition practices.

F. There are interventions that improve the health environment, such as improved access to safe drinking water and health services.

It is worth noting that many development interventions attempt to improve the broad environment in which households exist or to raise levels of human or physical capital. These do not directly affect food security outcomes. Instead, they raise incomes. One should not assume, however, that there is invariably a strong link between higher income and food security and nutrition outcomes.

In the case of nutritional status or food utilization, food is not the only input. Increased food access will not necessarily improve food utilization when other factors, such as the health environment, are not favorable. A second cause is ignorance. Households and individuals may simply not be aware of all the components of a healthy diet or of good health practices. The third reason for these weak links is that households, and individuals, face many competing demands for their limited financial resources. They may want to increase the level or quality of their food consumption, but they may also want to reduce labor drudgery, be better dressed, be able to send their children to school, and so on. In those projects that emphasize beneficiary participation, beneficiaries might choose interventions that have their largest impact on an outcome other than food security or nutrition.

An attraction of the framework here is that it provides some prior indications as to which interventions are most likely to have such an impact. For example, interventions directed at strengthening local institutions are unlikely to have a direct impact on nutritional status. Further, greater beneficiary involvement in project selection, design, and implementation may also result in interventions that do not address food security and nutrition concerns. Put another way, the principal concerns of beneficiaries may relate to objectives that differ from those of the project designer who seeks to improve food security and nutrition. Such observations do not necessarily invalidate approaches such as greater beneficiary participation, but do highlight the challenges associated with linking these to food security and nutrition.

It is also important to note that the strength of these links is not constant across all households within a given population. As many development practitioners are aware, women often face particularly severe constraints or have access to weaker productive assets. There is reasonable evidence to suggest that they devote a larger share of resources under their control to food security and nutrition objectives.

This provides the potential for a clear win-win scenario. Interventions directed toward women both relieve constraints on a particularly disadvantaged group and have maximal impact on food security and nutrition indicators.

Accordingly, an attraction of this conceptual framework is that it encourages project staff to consider carefully the likely impact of a proposed intervention on food security and nutrition. A second attraction is that it indicates that staff, when designing interventions, need to obtain and interpret information on the following questions:

- Who is food-insecure or at nutritional risk? Or, where should this intervention be located in order to maximize impact on these indicators?
- Why are they food-insecure or at risk? Or, what interventions will have maximal impact on improving these indicators?
- How best can this intervention be monitored and evaluated? Or, how can staff assess how well the project is working?

Table 1.1 Uses of this material at different points in the project cycle

			Points in the project cycle		
Chapter	Title	Brief description	Country background studies and project pre-planning	Formulation and appraisal	Implementation (monitoring and evaluation)
2.	Measuring nutritional dimensions of household food security	Outlines different measures of nutrition and explains how these can be implemented	✔	✔	✔
3.	Choosing outcome indicators of household food security	Outlines different measures of food security and explains how these can be implemented	✔	✔	✔
4.	Rapid appraisal techniques for the assessment, design, and evaluation of food security interventions	Outlines community-based methods for the assessment and monitoring of food security		✔	✔
5.	Constructing samples for characterizing household food security and monitoring and evaluating food security interventions	Reviews different methods of selecting a sample for needs assessment, monitoring, and evaluation		✔	✔
6.	Targeting: Principles and practice	Reviews different methods for targeting interventions		✔	✔
7.	Designing methods for monitoring and evaluating food security and nutrition interventions	Outlines rigorous, yet simple to implement, methods for project evaluation		✔	✔

Source: Compiled by author.

The next section introduces the material that provides answers to these questions.

Introduction to the Chapters

In addition to this introductory material, this book contains six chapters on different aspects of operationalizing food security and nutrition in development projects. Table 1.1 provides a list of these chapters and indicates where, within the project cycle, they can be used.

Some explanation of the particular topics chosen is warranted. Our focus on food security and nutrition reflects, in part, our own background and experience with development projects. But this does not mean that this book is only for practitioners in these fields. We hope that readers with a related interest, such as livelihood security, will also find this book useful. The material we present is the outcome of our interactions with project staff in multilateral and bilateral donor organizations, NGOs, officials in developing country governments, and project beneficiaries over the last three years. Working with these groups on project design and implementation helped improve our understanding of the largest gaps between theory and practice. This material attempts to fill these gaps. However, it does not pretend to be comprehensive. For example, although we discuss the design of survey instruments to obtain information on food security and nutrition (Chapters 2 and 3) and sampling methods for the implementation of such surveys (Chapter 6), we do not discuss the logistics of survey implementation because (1) our sense is that this is well-known territory for many development practitioners and (2) there are a number of excellent reference materials already in widespread circulation. Being selective rather than comprehensive also enabled us to write a shorter, and we think more manageable, volume with chapters that can be read either as stand-alone pieces or as a whole.

Set against these advantages are several disadvantages. First, we have no doubt missed some topics that at least one development practitioner would have included.

Second, the selective nature of this material might make it appear somewhat disjointed. However, we partly rectify this concern by noting that the chapters that follow can be grouped by their basic function in terms of assisting project staff in obtaining food security and nutrition and aiding in interpreting this information. Chapters that extensively discuss issues and techniques for obtaining information are Chapters 2 (nutritional dimensions of food security), 3 (choosing outcome indicators of food security), 4 (rapid appraisal techniques), and 5 (constructing samples). Chapters that emphasize the interpretation, use, and analysis of this information are 2 (nutritional dimensions of food security), 3 (choosing outcome indicators of food security), 6 (targeting), and 7 (designing methods for monitoring and evaluation).

Alternatively, these chapters can be grouped according to the questions, listed at the end of the previous section, that they answer. Specifically, the following chapters can be used to:

- identify who is food-insecure or at nutritional risk—chapters 2, 3, 4, and 5;
- identify causes of food insecurity and nutritional risk and the interventions that will alleviate these causes—this introduction, plus chapters 2 and 4;
- design monitoring and evaluation mechanisms—chapters 2, 3, 4, 5, 6, and 7.

The Chapters in Brief

Chapter 2: Measuring Nutritional Dimensions of Household Food Security

Many development projects intended to improve nutrition are constrained by a limited knowledge base. In particular, it is not clear whether the constraining factor to improved nutrition is poor access to food; weaknesses in the provision of health care, child care, or in the general health environment; or some combination of these. This chapter explains how such knowledge bases can be expanded using the principles of nutritional assessment. It answers the following questions: What is nutritional assessment? How can nutritional assessment assist the process of targeting projects to those most in need? How can nutritional assessment direct the selection and sequencing of interventions? How can nutritional assessment guide project monitoring and evaluation?

Chapter 3: Choosing Outcome Indicators of Household Food Security

Any commitment to improve food security and nutrition carries with it an important implication, namely the need to measure food security outcomes at household and individual levels. Measurement is necessary to characterize the severity of the food security problem and to provide a basis for measuring impact. This chapter shows how to construct four measures of household and individual food security: individual intakes, household caloric acquisition, dietary diversity, and coping indices. For each, an explanation is given regarding what this indicator measures, how the data is collected, and how indicators of food security are calculated. Each description ends with a commentary on the strengths and weaknesses of the method. This is followed by an explanation of how these different measures can be compared, illustrated using data collected in the Zone Lacustre region of Mali. The guide also proposes a possible sequence of activities that would use these indicators at different stages of the project cycle.

Chapter 4: Rapid Appraisal Techniques for the Assessment, Design, and Evaluation of Food Security Interventions

Participatory appraisal techniques are "a family of approaches and methods to enable rural people to share, enhance, and analyze their knowledge of life and conditions, to plan and to act" (Chambers 1994). These include mapping activities, transect walks, seasonal calendars, wealth ranking, and analytical diagramming. Unlike traditional, more extractive data-gathering methods, participatory rural appraisal (PRA) techniques are premised on the notion that local people have an enormous amount of local knowledge. Rather than merely appropriating this information, in PRA local people dominate the agenda, decide how to express and analyze information, and plan and evaluate.

This chapter outlines the advantages and disadvantages of rapid appraisal techniques in the context of food security interventions. These techniques are low-cost; provide information quickly; require little equipment; and by deliberately seeking local opinions, provide insights that might be missed by more conventional methods. But they require highly skilled personnel and are not suitable for targeting purposes. Six rapid-appraisal methods are outlined: concept definition; community mapping; household food security ratings; seasonal time lines; conceptual mapping of threats to food security; and the evaluation of interventions.

Chapter 5: Constructing Samples for Characterizing Household Food Security and Monitoring and Evaluating Food Security Interventions

Reliable information on household food security is a prerequisite for the accurate and effective design, monitoring, and evaluation of projects. But collecting data is not a costless exercise. This chapter discusses how random-sampling techniques—methods that use some mechanism involving chance to determine which farms, households, or individuals are to be studied—can economize on the costs of gathering information while increasing the likelihood that it will be both accurate and available in a timely fashion.

Chapter 6: Targeting: Principles and Practice

Many development agencies have a mandate to direct their investments toward the poor; that is, there is an explicit requirement that projects are targeted. This chapter introduces the principles underlying targeting, stressing that targeting only makes sense when the additional costs of doing so are outweighed by the additional benefits in terms of reduction in poverty or food insecurity. It also introduces the practice of targeting, beginning by distinguishing between two forms of targeting: administrative and self-targeting. Administrative targeting is the process by which project staff determine eligibility criteria. Under self-targeting, the intervention is, in principle, open to anyone who wishes to take part. However, it is designed in such a way that it is only attractive to certain households. The chapter explains how these methods can be implemented as well as their strengths and weaknesses.

Chapter 7: Designing Methods for Monitoring and Evaluating Food Security and Nutrition Interventions

In recent years, many development agencies have made intensive efforts to improve their efficiency and increase their impact on rural poverty. At the heart of this new strategic management process is the measurement of performance. But poorly thought-out evaluations may inadvertently act as an incentive to target better-off groups, which offer higher returns and promise faster disbursement of project resources. In addition, there is a clear danger of placing a higher priority on more easily measurable outcomes or indicators, which fail to provide the information necessary to address broader objectives or to enhance the effectiveness of rural development projects for "the poorest of the poor."

This chapter emphasizes the design of quantitative impact evaluation exercises for household food security and nutrition. It provides development practitioners with the basic principles on why, when, and how to choose and implement a particular evaluation system. Two key features of a good impact evaluation study are stressed: the availability of accurate baseline information and a properly thought-out control group, which allows before-after and with-without comparisons. The chapter also illustrates why the involvement of the evaluation team in the earliest stages of project design is the most suitable way to ensure a proper and accurate evaluation without having to rely on more complicated statistical techniques, as well as permit a sound learning process to ensue from the evaluation exercise.

2. Measuring Nutritional Dimensions of Household Food Security

Saul S. Morris

Development projects and practitioners can play a critical catalytic role in overcoming the nutrition problems of the rural poor, either by strengthening the household resource base for food and good health or by enhancing target groups' control and management of these resources.

Unfortunately, many practitioners do not have all the information they need to maximize the nutrition impact of rural development projects. This chapter outlines methodologies that will assist practitioners to improve the nutrition impact of development activities. The methodologies described here are jointly referred to as nutritional assessment. The chapter begins by explaining what is meant by nutritional assessment and how it can reinforce linkages between nutrition and agricultural development. It then considers how nutritional assessment can be used in rural development projects for beneficiary targeting and project formulation, as well as for practical project monitoring and evaluation.

Nutritional assessment has great potential for geographical targeting at little additional cost. In addition, it is also a useful input into project formulation. It is invaluable at the monitoring and evaluation stage because it offers the possibility of directly measuring the human-welfare impact of development activities, and also because the information generated cannot easily be manipulated by interested parties. In the final section of the chapter, the theoretical discussions are illustrated using data from Honduras.

Those interested in reading about these topics in more detail should consult Gibson (1990) and WHO (1995).

Background: The Role of Nutritional Assessment in Meeting the Challenge of Hunger and Poverty

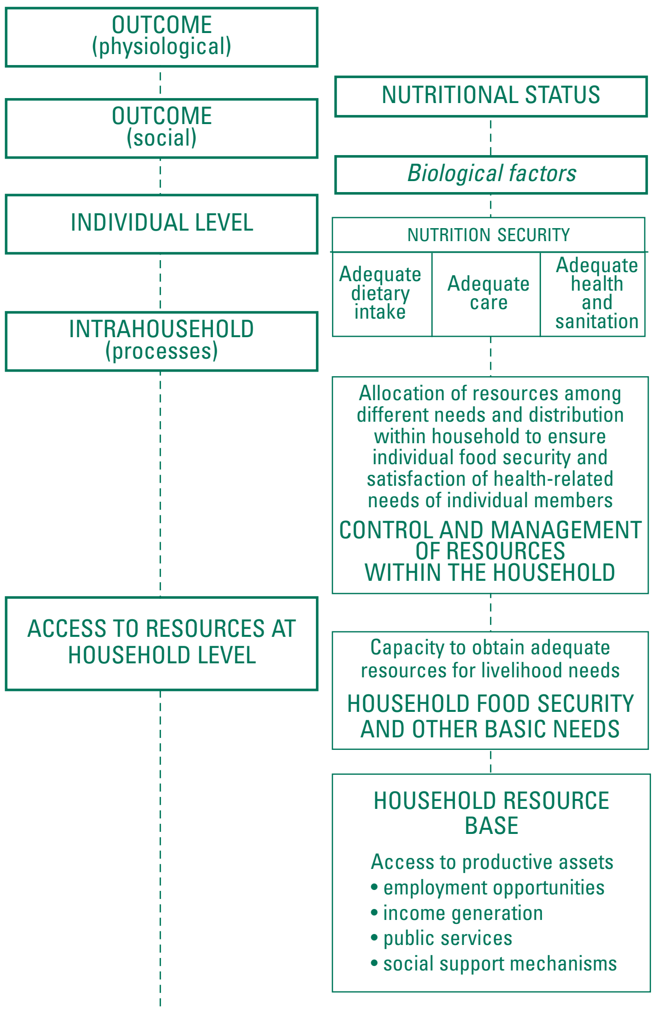
Nutritional assessments are measurements of body size, body composition, or body function intended to diagnose single or multiple nutrient deficiencies. Sometimes nutritional assessments consist of highly controlled technical measurements, while in other circumstances, they may be conducted in a participatory manner that fosters community involvement and ownership of the project as a whole. Findings may be interpreted at the level of the individual, but are commonly aggregated over a community, district, or subnational region.

Frankenberger et al. (1993) have shown that measures derived from nutritional assessments may be viewed as the biological manifestation of Nutrition Security, a "condition that combines having access to adequate food, being well cared for, and enjoying a healthy environment." The conceptual model developed by Frankenberger and colleagues is reproduced in Figure 2.1. In this model, rural development projects attempt to directly influence the household's resource base, and thus household food security.

Getting Familiar with Measures of Nutritional Status

There are numerous different measures of nutritional status, varying with respect to their ease of measurement, relation to dietary intake,

Figure 2.1 Nutritional security



Source: Frankenberger et al. 1993.

and velocity of change following shocks or improvements in the individual's environment. There are three major classes of measures. The first class uses clinical examinations to detect signs and symptoms of advanced nutritional depletion. Examples are surveys of goiter to detect iodine deficiency, or eye examinations to detect vitamin A deficiency. With appropriate training, lay inquirers can be used to determine levels of these conditions in the community.

In contrast, laboratory methods are usually invasive (involve taking samples from sites of the body that are not immediately accessible) and therefore poorly suited to routine use in a program situation. These methods are used to detect decreased levels of nutrients in body tissues or fluids, or decreased activity of an enzyme that is nutrient-dependent. One example of a laboratory method with potential for more general use is the detection of anemia by hemoglobinometry (see below).

The third class, and the main focus of this chapter, is anthropometry, or the measurement of body size and gross body composition. The basic principle of anthropometry is that prolonged or severe nutrient depletion eventually leads to retardation of linear (skeletal) growth in children and to loss of, or failure to accumulate, muscle mass and fat in both children and adults. These problems can be detected by measuring body dimensions, such as standing height or upper-arm circumference or total body mass (weight). All of these measures are expected to vary by the age and sex of the person measured, so that there is a need for the measurements to be standardized for age and sex before they can be interpreted.[1] Easy-to-use computer applications are available for these conversions.

The five most commonly used anthropometric indices are described in more detail in Table 2.1. There is a strong emphasis here on children under five years of age, because children are especially vulnerable to adverse environments and respond rapidly to changes. In particular, when children do not receive the nutrients they need, their growth is rapidly compromised, with long-term implications for their future productivity. On the other hand, although adults also lose weight in response to severe energy deficit, this effect can be very difficult to distinguish from their genetic potential. The selection of appropriate measures for different programmatic purposes is described in later sections of this guide.

Anthropometric measurements are subject to a number of sources of error, including instrument error, investigator error, and recall error (for measures based on age). These sources of error need to be controlled, since they can easily lead to overestimates of the frequency of malnutrition or underestimates of the effectiveness of interventions. Special standardization procedures have been developed to minimize measurement error (Habicht 1974).

Table 2.1 Commonly used anthropometric indices

Indicator	Age group	Requirements
Height-for-age/Length-for-age "Height" measured as recumbent length for under 2-year-olds. Measure referred to standard for well-nourished individual of same age and sex (usually National Center for Health Statistics [NCHS]).	Up to puberty	Extensive training required for measurement of recumbent length of infants and young children. Accurate age information required (often misreported in nonliterate societies).
Weight-for-age Measure referred to standard for well-nourished individual of the same age and sex (usually NCHS).	Up to puberty	Accurate age information required (often misreported in nonliterate societies).
Weight-for-height/Weight-for-length Weight measure referred to standard for well-nourished individual of same height and sex (usually NCHS). "Height" measured as recumbent length for under 2-year-olds.	Infancy and childhood	Extensive training required for measurement of recumbent length of infants and young children. Two different body measurements required.
Mid-upper arm circumference Special insertion tape used to identify midpoint of upper arm and measure circumference at this point.	All ages	Relatively little training required.
Body mass index Weight (kilograms) divided by height (meters) squared.	Adult	Two different body measurements required.

Source: Compiled by author.

Using Nutritional Assessment to Improve the Impact of Rural Development Projects

In the following sections, we show how nutritional assessment methods may be used to improve project formulation, beneficiary targeting, monitoring, and evaluation. Many of the approaches take advantage of the increase in the availability of nutritional data that has occurred since 1990 (see Chapter 4). Other approaches require collection of new data; interested readers are strongly advised to consult Chapter 6 before undertaking data collection activities.

Country Strategy, Project Inception, and Formulation

Currently available data on nutritional status may be especially valuable at the country strategy and project inception stages, both for targeting subnational regions and for needs assessment.

Where nutrition security is a priority, identifying the geographic areas of a country most in need of rural development interventions is facilitated by reference to existing sources of nutrition data. The principal nutritional indicator for targeting subnational regions is **proportion (absolute numbers) of children under 5, or of school age (6–10 years), with low height-for-age (stunting)**.

This indicator, more than any other, is recommended for identifying areas of greatest need for targeting economic and health interventions (WHO Expert Committee 1995). Weight-based measures are, in general, too sensitive to illness and specific childcare practices, and are subject to seasonal variations. Data on stunting are available at the subnational level for virtually all poor countries, from surveys (such as the Demographic and Health Surveys) and from school height censuses. The usefulness of the measure for project targeting can be enhanced by expressing the numbers on a per-kilometer-squared basis.

Regions/communities with large numbers of children characterized by low height-for-age are found in Africa, Asia, and Latin America and the Caribbean, though the condition is most common in South Asia (UN ACC/SCN 1998). In Latin America and the Caribbean, the prevalence of low weight-for-age (underweight) can be used as a proxy measure for low height-for-age (stunting), since the two indicators are highly correlated in this region where low weight-for-height (wasting) is not seen.

Another measure that may be of use in contexts of extreme poverty is proportion (absolute numbers) of adults and adolescents with low body mass index (BMI). This indicator identifies areas of severe food insecurity. Data are sometimes available at the subnational level from surveys (such as the Demographic and Health Surveys, which commonly assess the nutritional status of women of reproductive age). The usefulness of the measure for project targeting can be enhanced by expressing the numbers on a per-kilometer-squared basis. Caution should be exercised when using this measure in areas where advanced HIV disease is prevalent, since individuals with HIV disease are thin, but not as a result of food insecurity.

Regions/communities with large numbers of adults/adolescents characterized by low BMIs are found in Asia and Africa. Small mid-upper arm circumference is sometimes used as a proxy measure for low body mass. It should be noted that the presence of significant numbers of adults (say, 10 percent) with very low BMIs normally indicates a need for emergency relief rather than rehabilitation or development.

Needs Assessment

Also at the project formulation stage, nutritional measures can be reviewed to assess the needs of project beneficiaries. Normally this process is carried out for the project area as a whole, but where possible, it is informative to disaggregate by variables known to be linked to nutrition security, such as landownership, gender of household head, sanitary/health care resources, and so on. Nutrition indicators for needs assessment are described in Table 2.2.

The needs assessment process should start by collating nutritional data from as many different population-based sources as possible (data collected at health centers or from currently operating selective programs are much more difficult to interpret because of the inevitable biases).[2] The information should be arranged by indicator, age group studied, and year of collection. Conflicting evidence from different sources should be carefully reviewed with the help of local experts to identify the source of the discrepancy. Subsequently, it may be helpful to rank the different problems identified according to their frequency in the population.

It is useful to compare the same indicator across different age/sex

Table 2.2 Nutritional indicators for needs assessment exercises

Indicator	Interpretation
Prevalence of low height-for-age (stunting) in preschool or school-age children	Children's skeletal (linear) growth compromised due to constraints to one or more of nutrition, health, or mother-infant interactions. In some populations, these constraints are already apparent in utero. Quality of diet a more frequent limitation than inadequate quantity.
Prevalence of low weight-for-height (wasting) in preschool or school-age children	Children suffer thinness resulting from energy deficit and/or disease-induced poor appetite, malabsorption, or loss of nutrients.
Prevalence of low weight-for-age (underweight) in preschool or school-age children	This indicator confuses the two processes described above and is therefore not a good indicator for needs assessment purposes.
Prevalence of low body mass index (BMI) in adults or adolescents	Adults suffer thinness as a result of inadequate energy intake, an uncompensated increase in physical activity, or (severe) illness.
Prevalence of low mid-upper arm circumference in adults/adolescents	As above. Restricting analysis to the arm has the advantage of reflecting the mass of just three tissues—bone, muscle, and fat—the last two of which are particularly sensitive to body weight gain/loss.
Prevalence of low serum retinol in preschool children	Children suffer vitamin A deficiency, either as a result of low intake of vitamin A in the diet, or because there is a high frequency of infection, leading to sequestering of vitamin A from the blood.
Prevalence of low hemoglobin (anemia) in preschool or school-age children	Children suffer from anemia, either as a result of low iron intakes or poor absorption, or as a result of illness. Severe protein-energy malnutrition and vitamin B12/folate deficiency can also lead to anemia.
Prevalence of low hemoglobin (anemia) in nonlactating, non-pregnant women	Women suffer from anemia as a result of low iron intakes, poor absorption, illness, or excessive losses of blood. Severe protein-energy malnutrition and vitamin B12/folate deficiency can also lead to anemia.
Prevalence of low hemoglobin (anemia) in men	As above. Anemia is rare in adult men except in conditions of extreme iron-deficient diets.

Source: Compiled by author.

groups. For example, stunting in children where adults are also short may be more suggestive of intergenerational deprivation effects than of current food access or health problems; similarly, while wasting in children may indicate poor feeding practices or health problems, the combination of wasting in children and low body mass in adults indicates a crisis in entitlements to food. Specific nutrient deficiencies (for example, iron or vitamin A) are not uncommon in children as a result of poor feeding practices; however, when they are also found in adults, a problem of food access should be suspected.

It is also important to contrast different indicators. Substantial childhood wasting in the absence of stunting, for example, indicates a nutritional crisis of very recent advent. Stunting in the absence of wasting, on the other hand, indicates a complex and deep-rooted nutritional problem, sometimes not directly related to food availability at the household level. Similarly, specific nutrient deficiencies in the absence of stunting or wasting may indicate either poor feeding practices or a general problem of dietary quality, while, combined with stunting and/or wasting, they are more likely to indicate profound poverty of resources at many levels.

Project Implementation

Just as nutrition data can assist with targeting and needs assessment at the project formulation stage, so they can also be of assistance for small-area targeting and sequencing of interventions in the implementation phase.

The potential of small-area targeting is discussed in Chapter 6. This procedure is greatly facilitated when nutritional data are available at a fine level of disaggregation, permitting the identification of priority-need small areas (usually districts or municipalities) within the overall area of influence of the project.

School height censuses are an obvious source of such data, but detailed nutritional surveys are also occasionally available. Where such data are available, their use and interpretation are exactly as described in the previous section (see above). Where these data are not available, conducting a large-scale nutritional survey for the purpose of small-area targeting is likely to be cost-ineffective; other indicators should therefore be used (see Chapter 3).

Only in exceptional circumstances are nutritional data available at a level of disaggregation sufficiently fine to permit community-level targeting. Often, however, socioeconomic data can be collected that permit the estimation of the expected rate of malnutrition in the community.

For a number of reasons, it is unwise to use nutritional measures for household-level targeting in rural development projects. These reasons include the following:

- Many of the measures that have been discussed in the previous sections are dependent on the presence of a household member of a particular age and/or sex, and thus exclude a priori households of a different composition.
- Most nutritional measures are age-sensitive; for example, a two-year-old child is much more likely to be stunted than a one-year-old, even though the conditions of the household are identical.
- Some measures of nutritional status change in a relatively short time, so that a child who has just been ill can easily be wasted, even when the household's conditions are generally good.
- Many other measures reflect past conditions, or even intergenerational effects, more strongly than current conditions.
- The cutoffs used to determine the presence or absence of malnutrition are arbitrary, so that a child with a height-for-age Z–score of –2.1 is classified as stunted while one with a Z–score of –2.0 is not, even though there is little reason to include the first family in a development program and not the second.
- Finally, there have been instances where families in areas with projects using individual targeting-based nutritional status have actually withheld food from children so that their nutritional status will deteriorate and the family will be entitled to participate in the project.

The nutritional needs assessment described above is expected to identify the broad features of an appropriate nutrition strategy for the project area. Beyond this, the search for interventions should be guided by an analysis of the constraints to nutrition security in each of its contributing areas: household food security, health, and mother-infant interaction. As nutritional indicators represent the joint outcome of all of these factors, there is only a limited amount of information that they can provide on the causes of, and solutions to, nutrition insecurity.

Monitoring and Evaluation

Nutritional assessment can be an extremely valuable element of the monitoring and evaluation process in rural development projects for a number of different reasons:

- Nutritional measurements provide a measure of human welfare that is sensitive to changes in food supply, as well as to other community development processes.
- Nutritional measurements provide a nonsubjective, quantitative assessment of progress toward a fixed goal (the elimination of malnutrition).
- Nutritional measurements cannot easily be falsified by

individuals with vested interests in the outcome of the interventions (including the subjects themselves).

- Nutritional measurements are relatively easy to obtain, either in sentinel sites for the purpose of ongoing monitoring, or in a sample of the entire study area for the purpose of evaluation.

In order to assess whether project interventions have improved nutrition security among beneficiaries, it is first necessary to identify which nutritional indicators could plausibly have been altered by project interventions and which subgroups of the population are most likely to have benefited. For example, a project that has as its sole aim the promotion of home gardening should not be expected to produce an impact on adult BMI, since vegetables are, in general, rich in micronutrients but not in energy.[3] Similarly, a project aimed at increasing basic grain production in rural Africa is unlikely to affect the nutritional status of infants less than six months of age, since these infants usually consume only breast milk and are therefore unaffected by changes in the family diet. Relevant nutritional indicators for assessing the impact of a variety of different interventions are shown in Table 2.3.

The length of time that an intervention has been in place is also an important variable to take into account when selecting nutrition indicators and study populations, since different indicators reflect events in the recent and distant past with different intensities, and take different amounts of time to respond to such changes (Table 2.4).

Table 2.3 Nutrition indicators for monitoring and impact assessment

Intervention	Most relevant nutritional indicators
Improved availability of food (dietary energy) at the house-hold level, in areas where dietary energy intake is initially constrained.	Body mass index (BMI) (adults) Weight-for-height Z–score (two to five year olds) Weight-for-age Z–score (two to five year olds) Height-for-age Z–score (long-term evaluations only; two to five year olds)
Improved availability of food at the individual level, plus improvements in other basic needs, especially health	Height-for-age Z–score (under fives) Weight-for-age Z–score (under fives) Weight-for-height Z–score (under fives)
Increased intake of animal products	Anemia (hemoglobin) Serum vitamin A (retinol)
Increased intake of fruits and leaves	Serum vitamin A (retinol)

Source: Compiled by author.

Table 2.4 Time reference of different nutritional indicators

Indicator	Time reference for dietary influences
Serum vitamin A	Essentially, consumption over recent days, which can be influenced by consumption events up to four months in the past
Hemoglobin	Consumption over recent weeks and months
Weight-for-height, body mass index (BMI)	Consumption over recent weeks
Height-for-age	Cumulative life-time consumption, especially influenced by events occurring in first two years of life and prenatally
Weight-for-age	Mixture of weight-for-height and height-for-age effects

Source: Compiled by author.

There are many different ways of using nutritional assessment to determine whether project interventions are improving, or have improved, the nutrition security of the beneficiary population. In the following sections, we examine four such methods: (1) the use of sentinel sites for the monitoring of nutritional impact, (2) the examination of changes in nutritional status of populations before and after implementation of project activities, (3) the analysis of changes in the nutritional status of individuals before and after implementation of project activities, and (4) the comparison of achieved nutritional status across beneficiary and nonbeneficiary populations.

Sentinel sites for monitoring nutritional status

Sentinel sites (a few purposively selected "representative" locations where data collection and analysis activities are concentrated) have frequently played a major role in project monitoring activities. For project management, the advantage of setting up a sentinel site system is that a relatively small number of people can be intensively trained to provide needed information in a timely and systematized manner. On the other hand, there is always the danger that the sentinel sites selected may not be representative of the project area as a whole, and that the data collected may become more or less reliable over time as those charged with the data collection master the techniques or, alternatively, lose interest in the monitoring process.[4]

The most important element of a successful monitoring system is a mechanism for ensuring that the data are promptly collated and analyzed so that they can feed into decisionmaking processes without delay. Nutritional indicators should be selected on the basis of the simplicity of measurement: weight-for-age would be the indicator of choice in many communities, although mid-upper arm circumference may be as good or better in communities where acute or seasonal food shortages are known to occur and to result in fluctuations in body mass. The analysis of the data should focus on obtaining moving averages[5] that reflect important changes in nutritional status without being excessively dominated by short-term "blips." It is likely to be necessary to control for the effect of the aging of the study cohort over time, as this leads to apparent improvements in nutritional status that are—sadly—illusory. Samples of approximately 100 individuals are likely to be sufficient for the monitoring of trends over time, with measurements perhaps every two or three months. Ongoing monitoring may be linked to the evaluation strategies described in the following sections, but it is important to realize that it does not, in and of itself, provide evidence of any impact of project activities. Rather, it indicates that within the intervention area, changes are or are not occurring in the direction expected, and are or are not of the desired magnitude (see Chapter 7).

Evaluating changes in nutritional status of populations before and after implementation of project activities

One popular way of determining the impact of project activities on nutrition security is to conduct one survey prior to implementation and another at the end of the evaluation period, examining changes in the nutritional profile of the population over the two points in time. This type of evaluation is credible if it can be demonstrated that the population surveyed is the same at each period in time (for example, a representative sample of all adult women in the project zone of influence). It is not necessary for the individuals in the survey to be the same; indeed, often it is unavoidable that the individuals are different, such as when the nutritional status of children under five is measured before and after a five-year development project. The comparison may be strongly influenced by factors specific to the timing of the two surveys. This is particularly the case when nutritional measures are used that are sensitive to short- or medium-term fluctuations in intake (for example, serum vitamin A). It is less of a problem when using measures such as height-for-age Z–scores, which reflect cumulative influences over a substantial period of time.

When the beneficiary population alone is studied, the evaluation can determine whether the observed changes in nutritional status are of the expected direction and magnitude, but is unable to causally link program activities to observed changes. When a "control" group is also measured at the same time points as the intervention group (see Chapter 7), it is possible to infer whether changes in nutritional status appear to be more beneficial in the intervention group than in the control group. The before-after comparison is usually expressed as the change in mean values of the nutritional indicator, but can also be expressed as the relative (or absolute) change in the proportion of the population with values below some critical measure. The latter comparison may be more relevant from a human welfare perspective, but requires larger sample sizes than the comparison-of-means approach.

One factor specific to studies with nutritional status as outcomes is that the interpretation of the results will be strongly influenced by the age composition of the study population. If the age composition has changed between initial and final surveys, or if the intervention group has a different age structure from the control group, be it ever so slight, this must be taken into account in the analysis. Since adjusting for age effects requires some knowledge of statistical methods, the utmost care should be taken to ensure comparability of the initial and final samples.

Analysis of changes in nutritional status of individuals before and after implementation of project activities

In some situations, it is possible to track individuals over time and to examine associations between project activities and changes in nutritional status at the individual level. This approach to measuring project impact is expected to be far more sensitive than the approach outlined above.[6] An individual's final height minus initial height is referred to as gain in height, while their final weight minus initial weight is referred to as weight gain. Since the amount of gain in height and/or weight is dependent on the time elapsed between the two measures, it may be appropriate to express these measures as gain per unit time, usually referred to as height or weight velocity. It is very important to realize that height and weight velocity are both sex- and (especially) age-dependent, so that analysis must take account of different age structures of intervention and control groups. One other complication that should also be borne in mind is that many individuals will not be able to be traced at the time of the second survey. Since these individuals are always different from those who remain traceable, the picture of project impact obtained may be unrepresentative.

It is not a good idea to calculate an individual's change in Z–score from one time period to another, since, for example, a half Z–score deterioration in nutritional status in an infant can have very different physiological implications from a half Z–score deterioration in an older child. Such comparisons are also confounded by technical problems with the customarily used National Center for Health Statistics (NCHS) reference. A new NCHS reference has been available since December 2000.

Comparison of achieved nutritional status across beneficiary and nonbeneficiary communities

In the absence of data on nutritional status prior to intervention, it is possible to directly compare the attained nutritional status of children of project beneficiaries with the attained nutritional status of children of nonbeneficiaries. In order to be able to interpret the results of such a comparison, it is necessary either to assume that beneficiaries and nonbeneficiaries were comparable prior to the project intervention, or to adjust statistically for variables known to affect beneficiary status. The many dangers inherent to both approaches are explained in detail in Chapter 7.

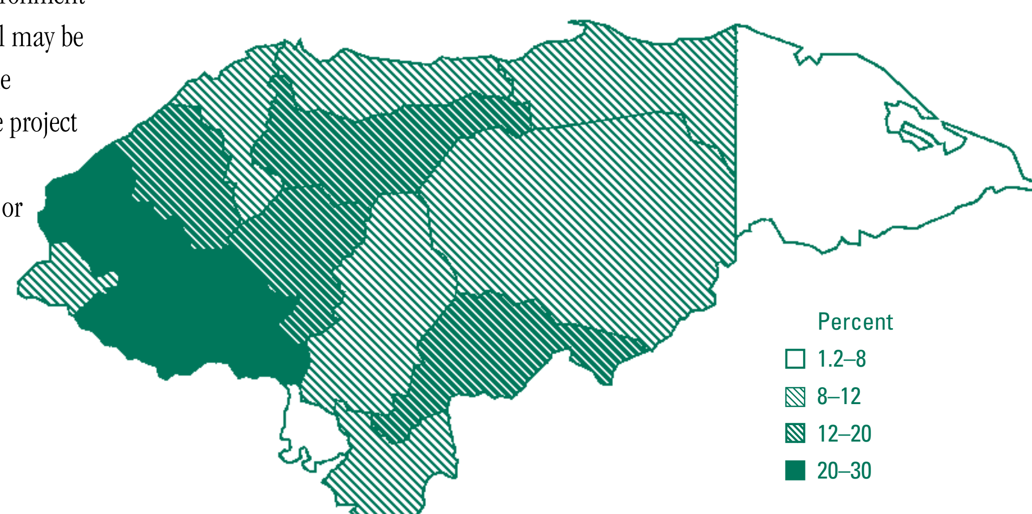
If all concerns about using these methods are satisfied, beneficiaries may be compared with nonbeneficiaries using either average (mean) nutritional status, or the proportions falling below a critical cutoff point. If the latter method is used, it is particularly important to select an indicator that can reasonably be expected to be sensitive to dietary intake and changes in the household environment over the period of evaluation. Some degree of internal control may be obtained by comparing the experience of two subgroups of the population. The first subgroup was expected to respond to the project interventions, while the second subgroup was not expected to respond to the particular kind of interventions implemented, or within the time frame under consideration.

Case Study of the Rural Development Plan for the Western Region, Honduras

Project Placement

Figure 2.2 shows the prevalence of severe stunting (height-for-age below –3 standard deviations of the NCHS median) in the 18 departamentos of Honduras. The data are taken from the Sixth Census of First-Graders' Heights (Republic of Honduras, Secretary for Education 1996). The prevalence of severe stunting exceeds 21 percent in four departamentos of the West (South-West) of Honduras: Copán, Intibucá, La Paz, and Lempira. The Rural Development Plan for the Western Region (PLANDERO) project[7] covers Copán and Lempira, but also Ocotepeque, where the prevalence of severe stunting is half that of Intibucá.

Figure 2.2 Percentage of severely stunted first graders, Honduras, 1996



Source: Compiled by author.
Note: Severe stunting is indicated by height-for-age Z–scores less than 3.

Figure 2.3 shows the number of severely stunted first graders per 100 hectares of land area. Intibucá, La Paz, and Lempira have the highest densities of malnourished children in the country, followed by Francisco Morazán and Cortés, areas where high population densities, rather than high prevalences of malnutrition, result in high concentrations of malnourished children. Copán is the sixth of 18

Figure 2.3 Density of severely stunted first graders per 100 hectares, Honduras, 1996

Source: Compiled by author.
Note: Severe stunting is indicated by height-for-age Z–scores less than 3.

departamentos when ranked by density of malnourished children, and Ocotepeque is the eleventh of 18. Table 2.5 shows the correspondence between the rankings based on the prevalence of malnutrition, and those based on the density per unit land area.

It appears that the location of the PLANDERO project is generally appropriate for a project aiming to affect nutrition security in Honduras, although it could be argued that it would have been preferable to exclude the *departamento* of Ocotepeque from the project's zone of influence.

Table 2.5 Severely stunted first graders per 100 hectares, and proportion of severely stunted first graders in the 18 *departamentos* of Honduras

Departamento	Severely stunted first graders per 100 hectares	Proportion of severely stunted first graders	
	(count)	*(percent)*	*(rank)*
Intibucá	1.83	30.70	1
Lempira	1.79	27.22	2
La Paz	1.56	21.11	4
Francisco Morazán	1.40	8.24	14
Cortés	1.32	7.95	15
Copán	1.23	21.27	3
Comayagua	1.05	13.77	7
Santa Barbara	.92	17.95	5
Yoro	.91	12.26	9
Valle	.84	7.79	16
Ocotepeque	.73	14.52	6
Atlantida	.67	8.51	12
Choluteca	.58	9.88	10
Colon	.50	8.46	13
El Paraiso	.45	12.80	8
Gracias a Dios	.34	4.49	17
Olancho	.29	9.52	11
Islas de la Bahia	.26	1.18	18

Source : Compiled by author from survey data.

Needs Assessment

Nutritional parameters for the project area are given in Table 2.6. The information has been collated from three different surveys and censuses conducted in recent years. Childhood stunting is the major nutritional problem in the area: The levels recorded, around 60 percent of all children, are among the highest in the world. There is virtually no wasting in this population, so that the relatively high levels of underweight can be attributed entirely to stunting. Similarly, there is very little chronic energy deficiency in adults: Although 8 percent of mothers of young children had BMIs below 18.5, virtually none had values below 17 (severe energy deficiency).

These sources reveal that in the area of influence of PLANDERO, the proportion of stunted preschool children rises from 33 percent in the highest (national) income quartile to 62 percent in the lowest quartile, and from 39 percent in households with high caloric adequacies to 67 percent in those with the lowest. The fact that stunting does not fall to low levels, even among those who are relatively well-off, may be attributed to (1) environmental features (for example, illness), which no one living in the region is protected from, and (2) intergenerational effects, reflecting the low stature of

Table 2.6 Nutritional indicators, western Honduras

Indicator	Age group	Percent detected	Geographical area	Source
Severe stunting (HAZ < –3)	First graders	22.6	Copán, Lempira, and Ocotepeque	Sixth Census of Height of First Graders, 1996
Stunting (HAZ < –2)	First graders	56.4	Copán, Lempira, and Ocotepeque	Sixth Census of Height of First Graders, 1996
Stunting (HAZ < –2)	Children < five years	60.0	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Household Consumption, Income, Expenditure, and Nutrition Survey 1994
Underweight (WAZ < –2)	Children < five years	32.8	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Household Consumption, Income, Expenditure, and Nutrition Survey 1994
Wasting (WHZ < –2)	Children < five years	3.5	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Household Consumption, Income, Expenditure, and Nutrition Survey 1994
Severe stunting (HAZ < –3)	Children 12–71 months	30.3	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Stunting (HAZ < –2)	Children 12–71 months	62.7	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Underweight (WAZ < –2)	Children 12–71 months	37.6	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Wasting (WHZ < –2)	Children 12–71 months	1.5	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Low body mass index (BMI< 18.5)	Mothers of children 12–71 months	8.3	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Low serum retinol (< 20 g/dl)	Children 12–71 months	18.7	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Anemia (Hemoglobin < 11 g/dl)	Children 12–71 months	29.7	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996
Anemia (Hemoglobin < 11 g/dl)	Mothers of children 12–71 months	26.7	Rural areas of Ocotepeque, La Paz, Lempira, and Intibucá	National Miconutrient Survey, 1996

Source: Compiled by author from survey data.
Note: HAZ stands for height-for-age Z–scores; WAZ stands for weight-for-age Z–scores.

the children's mothers (with an average height of only 148 centimeters) and growth retardation in utero.

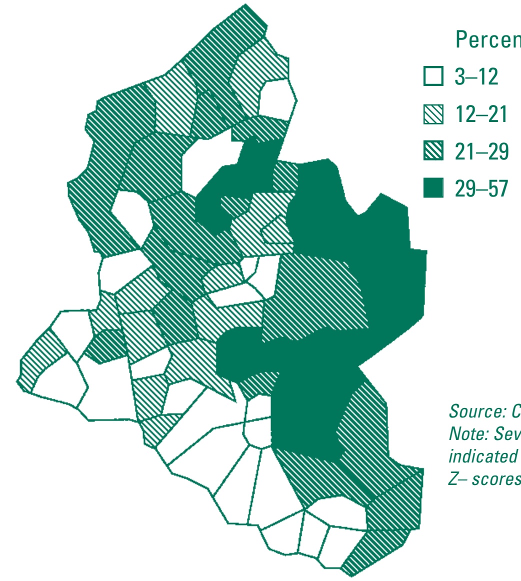
With respect to specific nutritional deficiencies, vitamin A deficiency (as measured by low serum retinol) constitutes a public health problem of "moderate" importance, according to international guidelines (WHO/UNICEF 1994). It is strongly associated with raised acute phase proteins (indicating infection), suggesting that it may result more from illness than from a lack of vitamin A in the diet per se (Republic of Honduras, Secretary for Health 1996). Anemia, on the other hand, is more common, both in children and in their mothers. There is also a strong association between anemia and infection, but the direction of causality cannot be determined.

These features suggest a population where ill health, poor care-giving practices, and perhaps dietary quality are likely to be major constraints to nutrition security, but an absolute deficit of dietary energy is not likely to be common. In these circumstances, increasing agricultural productivity alone cannot produce marked changes in nutrition security, even in the very long term. In order to affect nutrition security, PLANDERO might therefore choose to work in close coordination with health- and education-sector collaborators and invest in breaking down the isolation and poverty of the region in the longer term.

Targeting at the *Municipio* Level

Figure 2.4 shows the prevalence of severe stunting (height-for-age below –3 standard deviations of the NCHS median) in the 66 municipios of western Honduras. The data are taken from the Sixth Census of First-Graders' Heights (Republic of Honduras, Secretary for Education 1996). The prevalence of severe stunting exceeds 30

Figure 2.4 Percentage of severely stunted first graders, western Honduras, 1996



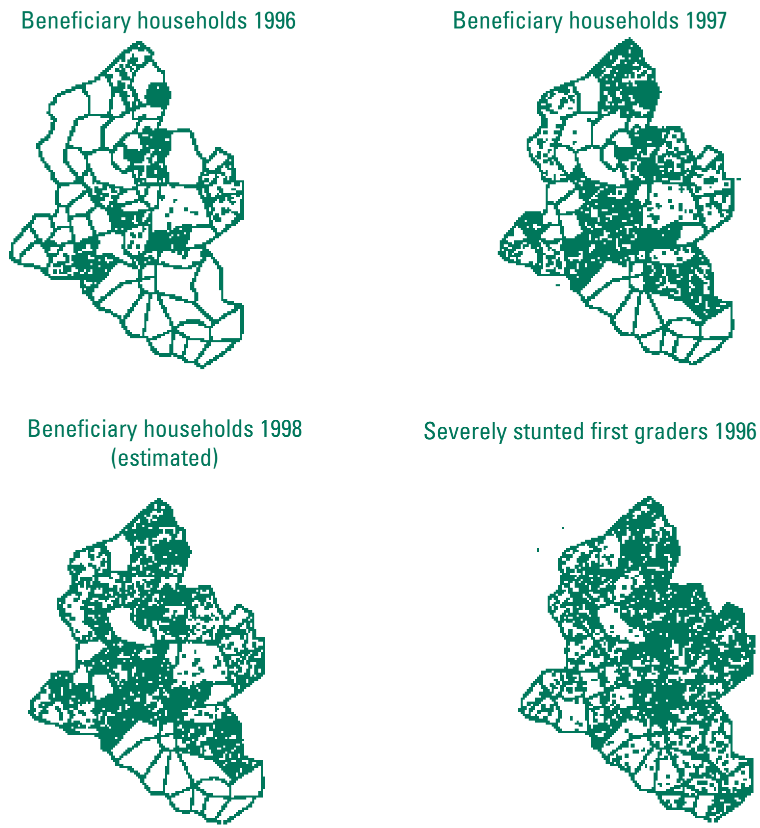
Source: Compiled by author.
Note: Severe stunting is indicated by height-for-age Z– scores less than 3.

percent in 13 *municipios* of the center, northeast, and northwest of Lempira, and the center-east of Copán, and is below 10 percent in 9 *municipios* of Ocotepeque, Southern Lempira, and the far south of Copán.

In order to assess the ability of PLANDERO to target its activities to the areas with the worst nutritional problems, each beneficiary family was given a score based on its *municipio* of residence. Families living in *municipios* with the highest levels of stunting were given the highest scores, while those living in the *municipios* with the lowest levels of stunting were given the lowest scores.[8]

The distribution of severely stunted first graders and of the project's beneficiary families in the first three project years (estimated numbers for 1988) is shown in Figure 2.5. The average stunting and severe stunting scores for beneficiary families in 1996, 1997, and

Figure 2.5 Distribution of PLANDERO beneficiary households and malnourished first graders, 1996–98, western Honduras



Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region. Each dot represents two households in the first three maps, and 2 first graders in the fourth map.

1998 (estimated) are shown in Table 2.7. The project was geographically neutral in its targeting in the first and second years of enrollment. In the third year (1998), new project beneficiaries were somewhat more likely to live in areas with more severe nutritional problems, so that the average scores of the new households were 63.5 (stunting) and 27.5 (severe stunting). However, the number of new

Table 2.7 Frequency of severe stunting among first graders, and severe stunting score of beneficiary households, western Honduras

Region/Program year	Number of individuals	Stunting score[a]	Severe stunting score
All first graders			
Western Honduras 1996	23,129	56.4	22.6
Program beneficiaries			
PLANDERO 1996	1,632	56.2	22.8
PLANDERO 1997	3,930	56.3	22.1
PLANDERO 1998[b]	5,109	57.9	23.4

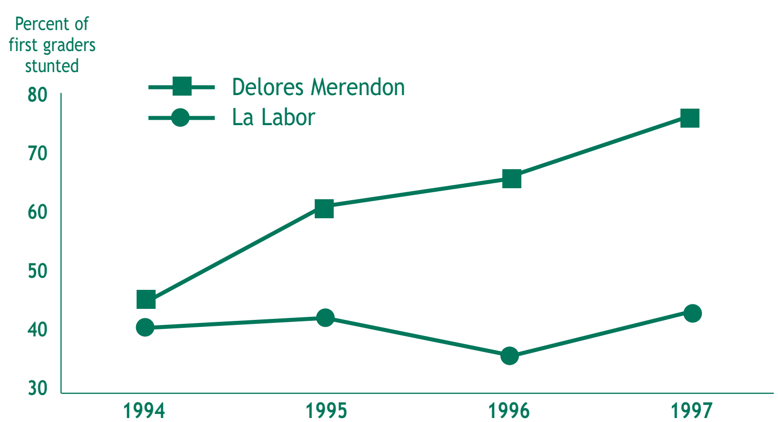
Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region.
a. See the discussion of stunting scores under Targeting at the Municipio Level on page 23.
b. Data for 1998 are estimates given no definitive data on the new 1998 beneficiary households was available at the time of writing.

beneficiary households anticipated in 1998 was small relative to the number already included in the project (30 percent of 1996 numbers), with the result that the overall targeting of project activities remained essentially neutral.

Monitoring

Figure 2.6 shows the proportion of first graders stunted for each year from 1994 to 1997 in two almost adjacent *municipios* in western Honduras. In the *municipio* of La Labor, which had a relatively strong institutional presence in 1993 (17 percent of farmers receiving technical assistance and 25 percent receiving credit) and was one of the first *municipios* to have groups assisted by PLANDERO in 1996, the rate of stunting remained almost unchanged throughout the period at approximately 40 percent. On the other hand, in Dolores Merendón, which had limited institutional presence in 1993 (7 percent coverage of technical assistance and 6 percent of farmers receiving credit) and did not receive any assistance from PLANDERO in 1996 or 1997, the rate of stunting in first graders increased dramatically from just over 45 percent in 1994 to nearly 75 percent in 1997. Although far from providing conclusive evidence of project impact, this example shows how it is possible to take advantage of existing data collection activities and extract potentially useful information about the evolution of nutritional status in the project zone of influence and in selected control areas. The analysis would have been strengthened if a finer level of geographical disaggregation could be achieved, making it possible to examine the experience of communities with a high coverage of project activities; alternatively, PLANDERO could have undertaken its own data collection activities in selected sentinel sites and compared the experience of its own study population with that of the universe of first graders in the same *municipios*.

Figure 2.6 Prevalence of stunting in two *municipios* of western Honduras, 1994–97



Source: Compiled by author from survey data.

Evaluation

Table 2.8 compares the nutritional status of children from birth to 60 months of age in July/August 1997, and the same group (plus new births) seven to nine months later in March/April 1998. Results are shown separately for children living in PLANDERO 96 households and for those living in PLANDERO 97 households. The seven-to-nine-month interval between the two survey rounds encompassed the 1997–98 growing and harvest season, during which time both sets of households received technical assistance and credit from PLANDERO. The control community households could not be included in this analysis because they were not assessed prior to the final survey round.

Table 2.8 Mean anthropometric status of children under five, by survey year and program status, western Honduras

Indicator/Year		PLANDERO 96			PLANDERO 97		
		(Z–scores)			*(Z–scores)*		
Height-for-age	1997	-2.09	(1.75)	n=243	-2.12	(1.69)	n=215
	1998	-1.99	(1.51)	n=245	-2.18	(1.53)	n=250
	Change 1997–98	+0.10	(0.15)	P=0.51	-0.06	(0.15)	P=0.69
Weight-for-height	1997	-0.17	(0.98)	n=243	-0.17	(1.17)	n=217
	1998	-0.07	(1.07)	n=243	-0.04	(0.97)	n=249
	Change 1997–98	+0.10	(0.09)	P=0.27	+0.13	(0.10)	P=0.19
Weight-for-age	1997	-1.39	(1.28)	n=243	-1.42	(1.32)	n=214
	1998	-1.29	(1.06)	n=243	-1.35	(1.13)	n=250
	Change 1997–98	+0.11	(0.11)	P=0.31	+0.06	(0.11)	P=0.57

Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region. Change is adjusted change in mean values; figures in brackets indicate standard deviation or, in the case of rows showing change, standard errors; n denotes sample number; and P denotes P–value.

The analysis shows that over this period, there was little change in anthropometric status either among PLANDERO 96 children or

PLANDERO 97 children, for any of the three indices examined. Furthermore, there was scarcely any evidence of differences between the experience of the two groups of children, although PLANDERO 96 children performed very slightly better than PLANDERO 97 children on the height-for-age indicator. In cases such as these, formal statistical hypothesis testing has little to add to the analysis.

One factor that should always be borne in mind when evaluating data in which the nutritional status of a given population or subpopulation is assessed on more than one occasion is the possibility of some change in the age structure of the population(s), which might invalidate uncontrolled comparisons. In the case of western Honduras, the average age of the children surveyed in the PLANDERO 96 communities was slightly different in July/August 1997 from that of the children in the PLANDERO 97 communities: 29.5 months versus 31.4 months, respectively. By March/April 1998, both study groups had aged somewhat, but this effect was more marked in the PLANDERO 97 communities, so that the average ages of children surveyed at this time were 33.5 months and 36.7 months, respectively. Changes in average anthropometric indices, adjusted statistically for changes in the age structure, are shown in Table 2.9.

Table 2.9 Change in anthropometric status of children under five between July/August 1997 and March/April 1998, adjusted for changes in the age structure of the survey populations, western Honduras

Indicator	PLANDERO 96			PLANDERO 97		
Height-for-age	+0.13	(0.14)	P = 0.36	+0.06	(0.15)	P = 0.67
Weight-for-height	+0.06	(0.09)	P = 0.52	+0.09	(0.10)	P = 0.37
Weight-for-age	+0.09	(0.10)	P = 0.38	+0.09	(0.11)	P = 0.41

Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region. Change is adjusted change in mean values; figures in brackets indicate standard deviation; and P denotes P-value.

This adjustment is sufficient to reverse the apparent direction of the evolution of height-for-age status in the PLANDERO 97 communities, so that their experience became comparable with that of the PLANDERO 96 communities. Thus, although technically demanding, age adjustment can be important to ensure the correct interpretation of results.

Many of the children in this data set were measured both in 1997 and in 1998, making it possible to examine changes at the individual level. Table 2.10 shows that when this approach is taken, there appears to be a rather substantial (and almost statistically significant, at the 5 percent level) difference in weight gain between children living in PLANDERO 96 communities and those living in PLANDERO 97 communities, in favor of the former. However, this difference is attenuated when differences in the age composition of the two groups are taken into account as described above. The approach that focuses on individual change has the advantage of not confusing the impact of changes in individual status with the impact of modifications in the composition of the group studied. On the other hand, it is marred by the (possibly major) biases inherent in studying only those children present in both surveys. Evaluators therefore need to carefully weigh the benefits and costs that would result from adopting this "cohort" approach.

In the absence of data on the anthropometric status of children in the control communities in 1997, any inference about the impact of PLANDERO on nutritional status relative to areas not included in the program must be drawn entirely from the *post*-intervention observations of March/April 1998. In order to extract the maximum possible amount of information from these data, it is useful to graph average height-for-age Z–scores recorded at this time by program status (PLANDERO 96, PLANDERO 97, and controls). Such a graph is

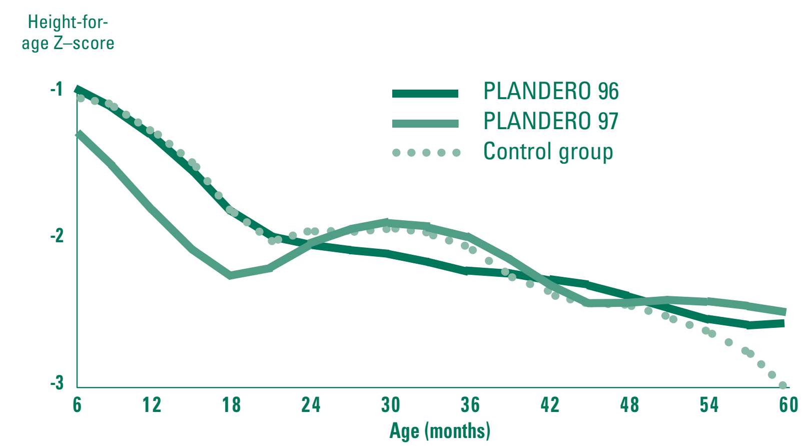
Table 2.10 Height and weight velocities of children under five living in the PLANDERO 96 and PLANDERO 97 study communities, Western Honduras, 1997–98

Indicator	Difference	PLANDERO 96	PLANDERO 97
		(centimeters per month)	
Height velocity		0.70 (0.42)	0.67 (0.36)
		n=179	n=178
	unadjusted difference	0.03 (0.04)	
		P=0.46	
	age-adjusted difference	-0.01 (0.03)	
		P=0.82	
		(grams per month)	
Weight velocity		193 (119)	169 (123)
		n=183	n=183
	unadjusted difference	24 (13)	
		P=0.055	
	age-adjusted difference	15 (11)	
		P=0.17	

Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region. Figures in brackets indicate standard deviation for height and weight velocity and standard error for differences; n denotes sample number; and P denotes P–value.

shown in Figure 2.7. Infants under six months of age were excluded because (1) they had been only briefly exposed to the project, and (2) most of their energy intake came from breast milk, which was unlikely to have been affected by the project activities. It can be seen that between six and 24 months of age, children from PLANDERO 97 communities were more stunted than PLANDERO 96 or control community children (who were similar to each other). From two years to 42 months, the PLANDERO 96 children were the ones to show most stunting. In this age group, the PLANDERO 97 and control community children had similar, higher height-for-age Z–scores. Above 48 months of age, the control community children were the most stunted. As stunting is basically established by two years of age in these communities, we can safely assume that the status of children four years and older represents the effects of variables that exerted their effect prior to the advent of PLANDERO in mid-1996. The experience of the younger children suggests a negative effect of PLANDERO's activities on stunting, but only in the year that each community first started to receive technical assistance and credit from the program. It can be seen that this analytic strategy is convenient when there are no baseline data available, but results are prone to the vagaries of sampling variation and interpretation can be somewhat subjective.

Figure 2.7 Average height-for-age Z–scores in March/April 1998, by program status



Source: Compiled by author from survey data.
Note: PLANDERO stands for the Rural Development Plan for the Western Region.

ENDNOTES

1. In order to allow for the normal variation in body size that is due to age and sex, observed measures are contrasted with the expected value for an individual of the same age and sex. For most commonly used anthropometric measures, these expected values are taken to be the average (median) value in the U.S. population, as determined by the National Center for Health Statistics (NCHS). The NCHS database is referred to as the reference population. How far above or below the reference median a particular value lies is measured in multiples of the standard deviation in the reference population, with the resulting quantity being referred to as a Z–score. Thus, Z–scores are calculated as

$$Z = \frac{\text{observed} - \mu_{NCHS}}{\sigma_{NCHS}},$$

   where µNCHS denotes the reference median, and –NCHS denotes the age-specific standard deviation in the reference population. There is currently much debate about the appropriateness of using the a single reference database to assess the growth of children and adolescents from different ethnic backgrounds, but it has generally been found that children from all countries and races can grow equally well up to 7 years of age (Habicht et al. 1974). At this age, height differentials within a race between children from different socioeconomic groups can reach 10 centimeters, while differences between races among children of high socioeconomic status do not exceed 1 centimeter.
2. The health profile of those attending health facilities is generally quite unrepresentative of the population as a whole, since people go to health facilities because they are sick. Similarly, those benefiting from selective programs are also unrepresentative, since such programs often target (or are self-targeted) towards the most needy. Alternatively, certain segments of the population may have characteristics that make it easy for them to access programs and services; such characteristics are likely to be associated with better health outcomes.
3. It is, of course, possible that the consumption of vegetables could displace other energy-rich items in adults' diets.
4. The intensive "training effect" may itself alter the nutritional status of the population, particularly if it is accompanied by increased awareness of nutritional issues.
5. Moving averages are averages of community-level nutritional status over a number of different time points (often up to five or more). These averages are recalculated every time measurements are made, so that short-term variations are "smoothed" by combining them with other measurements from different time-points. Medium-term trends are, however, reflected.
6. Why this should be the case may be understood by considering the case of an indicator such as height-for-age Z–score. At the initial survey, children's height reflects the sum of all environmental influences they have been exposed to since conception. On the other hand, their height at the final survey will reflect the sum of all the environmental influences they were exposed to from conception to project baseline, plus influences experienced during the course of project implementation. For the youngest children, the influences experienced during the implementation period will dominate the final measure; however, these children may have been buffered against external influences by their mothers. The final height of older children, on the other hand, will be dominated by events that occurred before the beginning of the project, and therefore is not particularly informative with respect to project impact.
7. See Report and Recommendation of the President to the Executive Board on a Proposed Loan to the Republic of Honduras for the Rural Development Plan for the Western Region (PLANDERO). International Fund for Agricultural Development 1993.
8. The scores assigned to each beneficiary household were equal to the rate of malnutrition in the *municipio* where the family resided. Thus, beneficiary families living in a *municipio* where 60 percent of all first graders were stunted were assigned a score of 60 each, while beneficiary families living in *municipios* where only 30 percent of first graders were stunted were assigned a score of just 30. The summary score for the whole project at any given point in time is calculated as the average of the scores assigned to each beneficiary household. The project may be described as neutral in its

geographical targeting if the average score thus derived is the same as the prevalence of stunting in the area as a whole. If, on the other hand, the average score is higher than the prevalence of stunting in the region, then the project is targeting areas with more severe nutritional problems; similarly, if the score is lower than the prevalence of stunting, then the project is targeting areas with less severe problems. The process was repeated using rates of severe stunting (height-for-age Z–score <–3) instead of rates of total stunting (Z–score <–2).

3. Choosing Outcome Indicators of Household Food Security

John Hoddinott

Introduction

Many development agencies consider household food security a guiding principle for designing interventions in rural areas. A commitment to food security—defined as the condition in which a population has the physical, social, and economic access to safe and nutritious food over a given period to meet dietary needs and preferences for an active life—carries with it an important implication for development practitioners, namely the need to measure food security outcomes at the household and individual levels. Measurement is necessary at the outset of any development project to identify the food-insecure, to assess the severity of their food shortfall, and to characterize the nature of their insecurity. Further, an initial measurement provides the basis for monitoring future progress and assessing the impact of these projects on the beneficiaries' food security.

The concept of food security has evolved considerably over time, as have food security indicators. There are approximately 200 definitions and 450 indicators of food security. One volume on household food security by Maxwell and Frankenberger (1992) lists 25 broadly defined indicators. Riely and Moock (1995) list 73 such indicators, somewhat more disaggregated than those found in Maxwell and Frankenberger. Chung et al. (1997) note that even a simple indicator such as a dependency ratio can come with many different permutations. They list some 450 indicators. With this abundance of indicators, an important methodological problem for development practitioners is to determine which indicators are appropriate, given the project being proposed.

Maxwell and Frankenberger (1992) make a distinction between "process indicators," which describe food supply and food access, and "outcome indicators," which describe food consumption. Many studies have found that process indicators are insufficient to characterize food security outcomes. Chung et al. (1997) found that there is little correlation between a large set of process indicators and measures of food security outcomes. This finding echoes the conclusion of some development agencies, namely that there is little correlation between area-level food production and household food security (IFAD 1997, 13).[1] For these reasons, this guide focuses only on outcome indicators.

The practical circumstances in the field are another factor that influence the choice of indicators. Development agencies and their local collaborators face significant financial and time constraints. Undertaking detailed household and individual surveys on an ongoing basis to characterize, monitor, and measure impact is not feasible, either because (1) the time spent on these activities does not fit into the standard project cycle, (2) the skills to implement and analyze such data are not available, or (3) purchasing these skills—by contracting to outside consultants, for example—is prohibitively costly. Mindful of this constraint, this guide shows how simple measures of food security outcomes can be constructed and compared. These methods are accessible to anyone with a basic grounding in statistics and access to a spreadsheet software program such as Microsoft Excel.

The next section outlines four ways of measuring household food security outcomes: (1) individual intakes, (2) household caloric acquisition, (3) dietary diversity, and (4) indices of household coping strategies.[2] In each case, there is a brief explanation of what this indicator measures, how data can be collected, and how indicators of food security can be calculated. Each description ends with a commentary on the strengths and weaknesses of the method.

It is possible that project designers may wish to use some combination of these indicators. For example, project goals might be specified in terms of improving caloric availability at the household level, yet there may not be sufficient resources to monitor this outcome on an ongoing basis. Section 3 explains how using simple measures of statistical association, together with these indicators, can overcome problems such as these. The final section proposes a possible sequence of activities that would use these indicators at different stages of a project cycle.

Readers of this chapter should note that the methods presented here are complemented by material in the introduction (on concepts of food security), in Chapter 2 (nutritional dimensions of food security), in Chapter 4 (on obtaining information on food security status using rapid appraisal techniques), and in Chapter 5 (sampling techniques for household surveys).

Outcome Measure of Household and Individual Food Security

This section outlines four ways of measuring household and individual food security: individual intakes (either directly measured or 24-hour recall), household caloric acquisition, dietary diversity, and indices of household coping strategies. This ordering of methods is deliberate, moving from methods that are time- and skill-intensive, but are regarded as being more accurate, to those that can be implemented quickly, are relatively undemanding in terms of the skills required for their implementation, but are more impressionistic.

Box 3.1 Energy content per 100 grams of edible portions, selected foods

Food	Kilocalories	Food	Kilocalories
Cereals and grains		**Grain legumes**	
Maize, yellow immature on cob	166	Beans/peas, fresh, shelled	104
Maize, white whole kernel, dried	345	Beans, dried	320
Maize, flour, 60–80 percent extraction	334	Chickpea, whole seeds, raw, dried	327
Maize meal	341	Cowpea, mature pods, dried	318
Millet, finger, flour	315	Mung bean, dried	322
Millet, bullrush, whole grain	339	Pigeon pea, dried	309
Rice, milled	333	**Nuts and seeds**	
Sorghum, whole grain	343	Bambara groundnut, fresh	346
Sorghum flour	337	Cashew nut, dried	560
Wheat flour	340	Coconut, mature kernel, fresh	392
White bread	240	Groundnut, dry	572
Brown bread	233	**Meat, poultry and eggs**	
Starchy roots, tubers		Beef, moderately fat	234
Cassava meal	318	Egg, hen	140
Plantain, ripe, raw	128	Goat, moderately fat	171
Sweet potato, raw	109	Mutton, moderately fat	257
Taro/cocoyam	94	Poultry	138
Yam, fresh	111	Fish, dried	255
Yam, flour	310	**Oils and fats**	
Sugars		Butter from cow's milk	699
Sugar	375	Coconut oil	900
Milk and milk products		Ghee, clarified butter	884
Milk, cow, whole	79	Lard/animal fats	891
Milk powder, cow, whole	357	Margarine	747
Milk, goat	84	Red palm oil	892

Source: CTA/ECSA 1987.

Individual Food Intake Data

Description. This is a measure of the amount of calories, or nutrients, consumed by an individual in a given time period, usually 24 hours.

Methods for generating these data. There are two basic approaches used to collect these data. The first is observational. An enumerator resides in the household throughout the entire day, measuring the amount of food served to each person. The amount of food prepared but not consumed ("plate waste") is also measured. The enumerator also notes the type and quantity of food eaten as snacks between meals as well as food consumed outside the household. The second method is recall. The enumerator interviews each household member regarding the food they consumed in the previous 24-hour period. This covers the type of food consumed, the amount consumed, food eaten as snacks, and meals outside the household.

Method of calculation. Data collected on quantities of food are expressed in terms of their caloric content, using factors that convert quantities of edible portions into calories. A useful reference point for these conversion factors is found on the Web site for the United States Department of Agriculture (USDA), http://www.nal.usda.gov/fnic/foodcomp, also available in hardcopy form (USDA 1999). Another source is CTA/ECSA (1987). A sample of these conversion factors is found in Box 3.1. These intake data are compared against a definition of food needs. It should be noted that "food needs" is a contested concept. Individual caloric requirements reflect individual characteristics such as age, sex, weight, body composition, disease states, genetic traits, pregnancy, and lactation status, and activity levels, as well as other factors such as climate. A typical approach is to begin with a reference person, say a 60-kilogram man aged somewhere between 30 and 60 years undertaking "moderate activity." This yields a caloric requirement of approximately 2,900 kilocalories per day. Individual requirements for children are made on the basis of their age and sex to yield "adult equivalents." These are reported in Box 3.2. A minimum requirement for a low-activity existence—8 hours sleeping, 1 hour walking, 15 hours standing or sitting quietly—is 2,030 kilocalories, or 70 percent of that required to undertake moderate activity. For this reason, this lower figure is often used as a cutoff to determine whether an individual is consuming enough to meet their food needs. However, it should be stressed again that there is no universal agreement on these figures; estimates of "basic requirements to meet food needs" range from 1,885 to 2,500 kilocalories (James and Schofield 1990; Smil 1994).

Box 3.2 Recommended daily caloric intakes

Age group	Kilocalories per day		
Young children			
<1	820		
1–2	1,150		
2–3	1,350		
3–5	1,550		
Older children	**Boys**	**Girls**	
5–7	1,850	1,750	
7–10	2,100	1,800	
10–12	2,200	1,950	
12–14	2,400	2,100	
14–16	2,650	2,150	
16–18	2,850	2,150	
Men	**Light activity**	**Moderate activity**	**Heavy activity**
18–30	2,600	3,000	3,550
30–60	2,500	2,900	3,400
>60	2,100	2,450	2,850
Women	**Light activity**	**Moderate activity**	**Heavy activity**
18–30	2,000	2,100	2,350
30–60	2,050	2,150	2,400
>60	1,850	1,950	2,150

Source: World Health Organization 1985.

Advantages and disadvantages of this method. This method has two principal advantages. First, when implemented correctly, it produces the most accurate measure of individual caloric intake (and other nutrients) and therefore the most accurate measure of an individual's food-security status. Second, because the data are collected on an individual basis, it is possible to determine whether food security status differs within the household. It may be that sufficient calories are being consumed at the household level, but inequalities within the household result in some members consuming in excess of their requirements, while others do not obtain sufficient food.

Set against these significant advantages are a large number of disadvantages. These measures of intakes need to be made repeatedly—ideally for seven nonconsecutive days—in order to account for within-person and within-household day-to-day variations in nutrient intake (for example, those resulting from religious prohibitions on the consumption of certain foods on certain days of the week or seasonal changes in diet). The method requires highly skilled enumerators who can observe and measure quantities quickly and accurately—and in a fashion that does not cause households to alter typical levels of food consumption and distribution within the household. The recall method requires enumerators to interview carefully every household member until they have established the exact makeup (food types, ingredients, and quantities) of every meal and snack, an extremely difficult task. This method generates an enormous amount of data that needs to be entered, checked, and aggregated before it is usable.

Household Caloric Acquisition

This is the number of calories, or nutrients, available for consumption by household members over a defined period of time.

Description. The principal person responsible for preparing meals is asked how much food was prepared for consumption over a period of time. After accounting for processing, this is turned into a measure of the calories available for consumption by the household.

Method for generating these data. A set of questions regarding food prepared for meals over a specified period of time, usually either 7 or 14 days, is asked to the person in the household most knowledgeable about this activity.[3] In constructing these questions, the following considerations should be borne in mind. First, it is extremely important that the list of foods specified in the questionnaire is detailed and exhaustive. Experience has shown that using short lists typically leads to an understatement of consumption by 25 to 75 percent (Deaton and Grosh 1998). Second, the questions need to unambiguously distinguish between the amount of food purchased, the amount prepared for consumption, and the amount of food served. Third, it is not uncommon for individuals to report consumption in units other than kilograms or liters. In such cases, it is necessary to obtain information on the size of a "heap," or the quantity contained in a "sawal," or whatever units are used locally.

Following is an excerpt from the questionnaire used in northern Mali to obtain information on food consumption in the last seven days.

We would like to ask you some questions about food consumption in this household in the last seven days. These questions pertain to the quantity of foods prepared for consumption.

Food	Unit	Quantity
Millet		
Sorghum		
Rice		
Maize		
Bread		

Note: Quantity consumed (units): 1."Bowl"; 2."Sack"; 3."Sawal"; 4."Pot"; 5."Calebash"; 6. Kilogram.

Method of calculation. Converting these data into calories requires three steps:

1. Converting all quantities into a common unit such as a kilogram.
2. Converting these into edible portions by adjusting for processing.
3. Converting these quantities into kilograms using the standard caloric conversions.

Sample data for five households consuming millet are reported below. Measurements undertaken as part of this survey determined that millet was typically measured in "sawal" and "pots." Both were obtained and the amount stored in these was weighed. One sawal contained 6 pots, and a pot was approximately 0.77 kilograms, implying that a sawal was 5 kilograms. The ratio of unground to processed millet, 0.61, was obtained by providing several women with 1 kilogram measures of millet, having the millet crushed using local technologies, and measuring what remained. The number of calories available in the previous seven days was computed by taking this quantity and multiplying it by the number of calories (3,390 kilocalories) in 1 kilogram of edible millet.

Advantages and disadvantages. This measure produces a crude estimate of the number of calories available for consumption in the household. It is not obvious to respondents how they could manipulate their answers. Because the questions are retrospective, rather than prospective, the possibility that individuals will change their behavior as a consequence of being observed is lessened. The level of skill required by enumerators is less than that needed to obtain information on individual intakes. In this locality, it took, on average, around 30 minutes per household to obtain these data, an amount of time considerably less than that required to obtain information on individual intakes.

Household	Quantity consumed				Number of calories available from consumption in the previous seven days
	Unit	Number	Conversion into kilograms	Adjustment for processing	
1	sawal	15	15 x 5 = 75	75 x 0.61 = 45.75	45.75 x 3,390 = 155,093
2	sawal	10	10 x 5 = 50	50 x 0.61 = 30.5	30.5 x 3,390 = 103,395
3	sawal	14	14 x 5 = 70	70 x 0.61 = 42.7	42.7 x 3,390 = 144,753
4	pot	12	12 x 0.77 = 9.24	9.24 x 0.61 = 5.63	5.63 x 3,390 = 19,086
5	pot	20	20 x 0.77 = 15.4	15.4 x 0.61 = 9.39	9.39 x 3,390 = 31,832

Set against these advantages are a number of disadvantages. This method generates a large quantity of numerical data that needs to be carefully checked both in the field and during data entry. Relative to the methods described below, data processing requirements are also higher. It is not as accurate as dietary intake data. The use of a recall period puts considerable reliance on recollection of events that may not be remembered accurately, with respondents either forgetting about particular foods or "telescoping"—including foods that had been used in a period prior to the preceding seven days. It is not especially accurate in capturing any food eaten outside of the household. It does not incorporate considerations of wastage, nor is it possible to uncover differential allocations of food among household members. Just as with the dietary intake method, it is necessary to convert quantities into calories and compare these against some standard, which, as already discussed, remains controversial.

Dietary Diversity

Description. This is the sum of the number of different foods consumed by an individual over a specified time period. It may be a simple arithmetic sum, the sum of the number of different food groups consumed, the sum of the number of different foods within a food group, or a weighted sum—where additional weight is given to the frequency by which different foods are consumed.

Method for generating these data. One or more persons within the household are asked about different items they have consumed in a specified period. Where it is suspected that there may be differences in food consumption among household members, these questions can be asked of different household members. Experience implementing this method has shown that comprehensive lists with 100 to 120 different food items perform better than shorter lists in distinguishing better-off from poorer households. Determining which items should appear on these lists can be done via rapid appraisal exercises (see Chapter 5), discussions with key informants, and references to previous survey work. Below is an excerpt from a questionnaire used in northern Mali, to illustrate this approach.[4]

I would like to ask you about all the different foods that you have eaten in the last 30 days. Could you please tell me whether you ate the following foods: 16 to 30 days in the last month—that is, at least every other day if not more frequently than that (J); 4 to 15 days in the last month—that is, once or twice a week (S); 1 to 3 days in the last month (M); 0 days—not at all (R).

	Frequency					Frequency			
Item	J	S	M	R	**Item**	J	S	M	R
Cereals					Fruits				
Millet					Bananas				
Sorghum					Mangoes				
Rice					Lemons				
Maize					Pineapple				
Bread					Other fruits				
Wheat					Meat				
Other cereals					Beef				
Tubers					Chicken				
Sweet potato					Sheep/goat				
Manioc					Fish				
Groundnuts					Dried				
Other tubers					Smoked				
Vegetables					Milk products				
Tomatoes					Cows milk				
Onions					Goats milk				
Beans					Other items				
Carrots					Butter				
Okra					Tea				
Other vegetables					Salt				

Methods of Calculation

There are two possible methods of calculation: (1) calculating a simple sum of the number of different foods eaten by that person over the specified time period, and (2) calculating a weighted sum, where the weights reflect the frequency of consumption and not merely the number of different foods. Here, the following weights are assigned: foods consumed at least every other day, if not more frequently: 24; foods eaten once or twice a week: 10; foods eaten infrequently (1–3 times per month): 3; and foods never eaten: 0.

Sample data for five households, together with these two measures, are presented below.

Household	Millet	Sorghum	Rice	Beef	Salt	Tea	Simple sum	Weighted sum
1	J	J	R	M	J	J	5	99
2	J	J	M	M	R	S	5	64
3	S	R	J	R	R	R	2	34
4	S	R	R	R	S	R	2	20
5	J	R	R	R	M	J	3	51

Advantages and disadvantages. The use of this measure stems from the observation made in many parts of the developing world that as households become better-off, they consume a wider variety of foods. It is easy to train enumerators to ask these questions. In addition, individuals generally find them easy questions to answer. Asking these questions typically takes about 10 minutes per respondent. Field-testing indicates that this measure is correlated with levels of caloric acquisition; tracks seasonal changes in food security (measures of dietary diversity are highest just after harvest time and lowest during the hungry season); and also appears to capture differences in distribution within the household. In northern Mali, for example, women reported that they were more likely than their husbands to reduce their own food consumption during periods of stress, and this was reflected in lower scores for women than for men on measures of dietary diversity. Finally, a diverse diet is a valid welfare outcome in its own right—the nutritional literature is placing increasing emphasis on the importance of consuming a wide variety of foods to enhance dietary quality in addition to addressing longer-standing concerns regarding quantities of consumption.

The disadvantage of this measure is that simple form does not record quantities. If it is not possible to ask about frequency of consumption of particular quantities, it is not possible to estimate the extent to which diets are inadequate in terms of caloric availability (but see footnote 4).

Indices of Household Coping Strategies

Description. This is an index based on how households adapt to the presence or threat of food shortages. The person within the household who has primary responsibility for preparing and serving meals is asked a series of questions regarding how households are responding to food shortages. In the nutrition literature, these first appeared in Radimer, Olson, and Campbell (1990). Coping strategies themselves are discussed in Maxwell and Frankenberger (1992). Maxwell (1996) proposed a method for taking consumption-related strategies and constructing a numerical index.

Method for generating these data. The most knowledgeable woman in the household regarding food preparation and distribution within the household is asked a series of questions of the following form.

In the last seven days:

1. Has the household consumed less preferred foods? (Circle the best response.)
   1. Never 2. Rarely (once)
   3. From time to time (2 or 3 times) 4. Often (5 or more times)

2. Have you reduced the quantity of food served to men in this household?
   1. Never 2. Rarely (once)
   3. From time to time (2 or 3 times) 4. Often (5 or more times)

3. Have you reduced your own consumption of food?
   1. Never 2. Rarely (once)
   3. From time to time (2 or 3 times) 4. Often (5 or more times)

4. Have you reduced the quantity of food served to children in this household in the last seven days?
   1. Never 2. Rarely (once)
   3. From time to time (2 or 3 times) 4. Often (5 or more times)

5. Have members of this household skipped meals in the last seven days?
   2. Never 2. Rarely (once)
   3. From time to time (2 or 3 times) 4. Often (5 or more times)

6. Have members of this household skipped meals for a whole day?

Method of calculation. A sample of responses to these questions, taken from a survey of households in the Zone Lacustre region of Mali, are reproduced below.

	Question					
Household	#1	#2	#3	#4	#5	#6
1	3	3	3	3	1	1
2	3	3	3	3	2	2
3	2	2	3	2	2	2
4	3	3	4	3	3	3
5	2	1	2	2	1	1

There are several ways of summarizing the information obtained from these questionnaires into a single number.

- Counting the number of different coping strategies used by the household. Here, this is the number of strategies that the household used often, from time to time, or rarely. The higher the sum, the more food-insecure the household.
- Calculating a weighted sum of these different coping strategies, where the weights reflect the frequency of use by the household. A simple way of doing this is to make the weights consecutive, so that "often" is counted as a 4, "from time to time" is counted as a 3, "rarely" is counted as a 2, and "never" is counted as a 1. The higher the sum, the more food-insecure the household.
- Calculating a weighted sum of these different coping strategies, where the weights reflect the frequency of use—as described above—and the severity of the household's response. A simple way of doing this is to ascribe a weight of 1 to the use of strategies such as eating less preferred foods (question 1) and reducing portion sizes served to men, children, and women (questions 2, 3, and 4); a weight of 2 to skipping meals

(question 5); and a weight of 3 to skipping eating all day (question 6). The first household on this list would obtain a score of $17 = 1 \times (3 + 3 + 3 + 3) + 2 \times (1) + 3 \times (1)$. Again, the higher the sum, the more food-insecure the household.

Household	Number of different strategies used	Weighted sum reflecting frequency of use	Weighted sum reflecting frequency and severity of use
1	4	14	17
2	6	16	22
3	6	13	19
4	6	19	28
5	3	9	12

Advantages and disadvantages of this measure. There are three attractive features of this measure. First, it is easy to implement, typically taking less than three minutes per household. Second, it directly captures notions of adequacy and vulnerability (is there enough food to eat in this household?), as well as the vulnerability of households. Those households, using a larger number of coping strategies or using more severe strategies, are more likely to be poor and more vulnerable to destitution. Third, the questions asked are easy to understand, both by respondents and by analysts and project designers.

There are also several disadvantages. As it is a subjective measure, with different people having different ideas as to what is meant by "eating smaller portions," comparison across households or localities is problematic. In particular, as part of the field tests for these measures, men and women were asked what constituted a "food-secure" diet. Poorer households tend to report smaller quantities of food than richer households. This has two implications. First, this measure can be somewhat misleading—a richer and a poorer household may both report eating smaller quantities, but this does not imply an equal increase in food insecurity. Second, evaluating the impact of an intervention solely in terms of this measure risks setting a lower target for poorer households than for richer ones.

Second, this measure's simplicity makes it relatively easy to misreport a household's circumstances. For example, households might perceive that they are more likely to receive assistance when they report greater use of these coping strategies. Finally, it is necessary to decide what weights should be applied to different questions and to different levels of response. The rapid appraisal techniques described in Chapter 5 could be used to obtain this information.

A Comparison of Methods

Table 3.1 provides a summary table that qualitatively compares these four methods in terms of costs, time and skill requirements, and susceptibility to misreporting.

Table 3.1 Comparison of methods in terms of costs, time, and skill requirements, and susceptibility to misreporting

Method Details	Individual intake	Household caloric acquisition	Dietary diversity	Index of coping strategies
Data collection costs	High	Moderate	Low	Low
Time required for analysis	High	Moderate	Low	Low
Skill level required	High	Moderately high	Moderately low	Low
Susceptibility to misreporting	Low	Moderate	Low	High

Source: Compiled by author from survey data.

Exploring Associations between Different Outcome Measures of Food Security

Each of the four measures described above are valid indicators of different dimensions of food security. However, there may be occasions when project designers, managers, or evaluators want to compare these indicators. For example, suppose that a project objective includes increasing levels of caloric availability at the household level, but there are insufficient financial resources to monitor this outcome on an ongoing basis. In such a circumstance, it would be useful to know whether changes in, say, dietary diversity are associated with increases in household caloric availability. Comparing these indicators may also provide insights into the distribution of project benefits within the household. For example, a finding that household caloric availability is rising, yet information on coping strategies indicates that women, but not men, are continuing to reduce food consumption during periods of stress, would be consistent with a project providing benefits, but these are being accrued primarily by men within the household.

Comparing these indicators in the manner described here requires the use of statistical techniques that measure the strength of association, or correlation, between these indicators. Below, three methods are discussed: correlation coefficients, contingency tables, and regression-prediction methods. All are illustrated using data collected in a project in northern Mali. The techniques are a little more technically demanding than the material presented in the previous section, but only a little. They can be implemented by anyone who has competently completed a basic undergraduate course (not degree) in statistics and has access to a spreadsheet computer package such as Microsoft Excel.

Correlation Coefficients

A simple approach to examining the validity of alternative measures of food security is to calculate measures of correlation such as Pearson or Spearman correlation coefficients. These are index numbers that show to what extent two variables are linearly related. They can take on values that range from –1 to 1. A priori, it is expected that the dietary diversity index and per capita calorie consumption are positively related, that is, both increase in value together. By contrast, the indices of coping strategies and per capita caloric availability should be negatively related. One would expect that increased reliance on coping strategies would be associated with lower food availability.

Examples are reported in Table 3.2. The measure of dietary diversity is the weighted measure based on data provided by women in these Malian households. The index of coping strategies is doubly weighted by the frequency of use of this strategy and the severity of the strategy.

Table 3.2 Pearson and Spearman correlation coefficient between caloric availability and two alternatives

	Correlation between calories available per person and	
Measure	**Weighted female dietary diversity**	**Doubly weighted coping strategy index**
Pearson	0.17**	–0.17**
Spearman	0.22**	–0.17**

Source: Compiled by author from survey data.
*Note: ** denotes statistically significant at the 1 percent level.*

Note that prior expectations are borne out: there is a positive correlation between dietary diversity and caloric availability and a negative correlation between the coping index and caloric availability. All four correlation coefficients are statistically significant at the 1 percent level. A more difficult question is how to interpret the magnitude of these coefficients, which are all roughly the same. It would appear that there is little to choose between these two measures. Both provide some correlation with the benchmark, but not at an especially high level.

A drawback to the use of correlation coefficients is that the correlation could be driven by just one part of the distribution of joint variables. Suppose that for most households, there is little correlation between dietary diversity and calorie consumption, but for very rich households, the correlation is quite high. As a consequence, the calculated coefficient might prove to be statistically significant. An additional problem is that of false correlation where some other variable is correlated with both measures, producing a false correlation between the two variables that are observed. A reasonable conclusion, therefore, is that these correlation coefficients are a good exploratory tool, but should not be the only method used.

Contingency Tables

Contingency tables cross-classify two variables by two or more attributes. In the tables below, households are classified by whether per-person caloric availability is above or below 2,030 kilocalories per-person per-day. Approximately one-third of households did not have access to even this minimal amount of food. Households are separately ranked by the alternative indicators and grouped according to whether or not they are in the bottom tercile for that ranking. Within these tables, there are three numbers of interest: specificity, the fraction of food-insecure households also classified by the alternative as food-insecure; sensitivity, the fraction of food-secure households also classified by the alternative indicator as food-secure; and a chi-squared test of whether there is a statistically significant association between these attributes.

Table 3.3a Contingency table of caloric availability and weighted dietary diversity

Attribute	Household is in bottom tercile when ranked by dietary diversity	Household is not in bottom tercile when ranked by dietary diversity	Total
Per capita caloric availability < 2,030 kilocalories	45	48	93
Per capita caloric availability > 2,030 kilocalories	39	134	173
Totals	84	182	266

Specificity: 45/93 = 0.48

Sensitivity: 134/173 = 0.77

Chi-squared test = 18.70**

Source: Compiled by author from survey data.
*Note: ** denotes statistically significant at the 1 percent level.*

Table 3.3b Contingency table of caloric availability and weighted coping strategy index

Attribute	Household is in bottom tercile when ranked by coping strategy index	Household is not in bottom tercile when ranked by coping strategy index	Total
Per capita caloric availability < 2,030 kilocalories	26	67	93
Per capita caloric availability > 2,030 kilocalories	80	93	173
Totals	106	160	266

Specificity: 26/93 = 0.28

Sensitivity: 93/173 = 0.54

Chi-squared test = 8.44**

Source: Compiled by author from survey data.
*Note: ** denotes statistically significant at the 1 percent level.*

These contingency tables indicate that there is a statistically significant correlation between these attributes. The dietary diversity measure performs better than the index of coping strategies in identifying food-secure and -insecure households as measured by caloric availability. This can be seen when comparing the measures of specificity and sensitivity in Tables 3.3a and 3.3b.

There are, however, two problems associated with using contingency tables. First, there is the issue of choosing the cutoffs for the attributes. Second, the method becomes demanding in a statistical sense when more than a handful of alternatives are considered. Specifically, repeating the exercise several times increases the likelihood of obtaining a significant association that results purely by chance. This can be rectified by setting a higher critical level for the chi-squared statistic (see Chung et al. 1997).

Regression-Prediction Methods

In light of the difficulties associated with correlation coefficients and contingency tables, a third method is outlined here that combines their advantages while minimizing their drawbacks. There is no formal name for this approach, which is described here as the "regression-prediction" method.

We begin with the observation that the two methods described above do not use all the information available. Specifically, in order to calculate per capita caloric consumption, it is necessary to determine how many people are in the household. Additionally, the location of the household will also be known. Consequently, these data can be added to the analysis at no additional cost. Further, there are good reasons for using this information. First, cross-country studies consistently reveal a negative association between food access and household size, although the reasons for this are not well understood (Deaton and Paxson 1998). Second, consider the following case. There are two localities: one is centrally located with a weekly food market; the second is remote from any markets. One would expect that the more remote village would face higher food prices and less access to a variety of foods. Failing to control for this might lead to a misleadingly strong association between dietary diversity and caloric consumption. The obvious way of incorporating these variables is to use them in a regression where the benchmark indicator is the dependent variable, and household size, location, and the alternative indicator appear as right-hand-side or explanatory variables.

Table 3.4 The relationship between (log) per capita caloric acquisition and two alternative measures of food security, controlling for (log) household size and location

Variable	Coefficient	t statistic	Coefficient	t statistic
Log household size	–0.403	6.338**	–0.339	5.338**
Dietary diversity	0.002	4.071**	–	–
Coping strategies	–	–	–0.053	1.764
Location				
Tomba	0.045	0.300	0.048	0.308
Mangourou	0.299	1.861	0.229	1.398
Gouaty	0.165	0.738	–0.140	0.656
N'goro	0.115	0.830	0.059	0.422
Tomi	0.092	0.467	–0.040	0.202
Hamakoira	–0.154	–0.872	–0.242	–1.345
Goundam Touskel	0.155	0.836	0.171	0.895
Ouaki	0.286	2.028	0.234	1.621
Anguira	–0.212	–1.283	–0.329	–1.976*
Constant	5.567	8.017**	8.495	42.885**
Adjusted R–squared		0.24		0.17

Source: Compiled by author from survey data.
*Note: ** denotes statistically significant at the 1 percent level; * denotes statistically significant at the 5 percent level; – indicates not applicable.*

The results of using this method for the Mali data, collected at the height of the hungry season, are presented in Table 3.4 (note that the dependent variable and household size have been transformed into their logarithmic values).

Controlling for household size and location, increased dietary diversity is associated with higher per capita caloric availability. Every additional point on the dietary diversity index is associated with an increase of 0.2 percent in caloric availability. This association is statistically significant at the 1 percent level. By contrast, once these other factors are taken into account, there is no statistically significant association between the coping index and the benchmark. Also note that the adjusted R–squared, which indicates to what extent the variance in the dependent variable is explained by the regression, is considerably higher for the regression using dietary diversity as an explanatory variable.

These estimated coefficients can be used to predict levels of log per person caloric availability. For example, for households residing in the village of Tomba, these predicted levels are

predicted log caloric availability per person

= 5.567 + 0.045 - 0.403 - log hh size

+ 0.002 - dietary diversity.

Taking antilogs yields predicted values in terms of caloric availability per person. These can be used to construct the following contingency tables in which the benchmark (actual caloric availability) and predicted caloric availability are divided into three categories: less than 2,030 calories per day (indicating severe food insecurity); 2,030 to 2,900 calories per day (indicating some degree of food insecurity); and greater than 2,900 calories per day. The results of this exercise for the Mali data set are reported in Tables 3.5a and 3.5b, with summary statistics reported in Table 3.5c.

Table 3.5a Contingency table of actual and predicted per-person caloric availability (dietary diversity)

		Number of households by predicted per-person daily caloric availability			Total
		< 2,030	2,030–2,900	> 2,900	
Number of households by actual per person daily caloric availability	< 2,030	50	34	9	93
	2,030–2,900	16	25	23	64
	> 2,900	12	39	58	109
Total		78	98	90	266

Table 3.5b Contingency table of actual and predicted per-person caloric availability (coping strategies)

		Number of households by predicted per person daily caloric availability			Total
		< 2,030	2,030–2,900	> 2,900	
Number of households by actual per person daily caloric availability	< 2,030	46	33	14	93
	2,030–2,900	12	34	18	64
	> 2,900	10	47	53	110
Total		68	114	85	267

Table 3.5c Comparison of predictive power of dietary diversity and coping index

Attribute	Dietary diversity	Coping index
Chi-squared test of association	60.16**	54.24**
	(percent)	
Households correctly categorized	50.0	50.0
Severely food-insecure households classified as food secure	9.7	15.1
Predicted distribution of calories per person by food-security status:		
< 2,030 (actual distribution = 35 percent)	29.0	25.0
2,030–2,900 (actual distribution = 24 percent)	37.0	43.0
> 2,900 (actual = distribution 41 percent)	34.0	32.0

Source: Compiled by author from survey data.
*Note: ** denotes significant at the 1 percent level.*
In Table 3.5a, household size, location, and dietary diversity were used as regressors;
in Table 3.5b, household size, location, and measure of coping strategies were used as regressors.

The chi-squared tests indicate that the match between the actual distribution of food acquisition and that predicted by both alternative indicators is greater than would have occurred if these alternatives had randomly assigned households to these different groups. Both correctly classify about half the households in the sample. Whereas the actual distribution across food security status is fairly constant, both alternatives predict that it is more concentrated among households experiencing moderate food insecurity. This is particularly marked in the case of the coping strategies index, which appears to especially underreport the number of severely food-insecure households.

Summary

This section has presented three methods for examining the associations among different outcome measures of food security. All three can be implemented using a standard spreadsheet package.

Developing and Using Outcome Indicators of Household Food Security in Development Projects

The material presented thus far has outlined possible outcome measures of food security and methods for evaluating these. This section outlines a possible sequence of events by which project designers can implement these methods. We are assuming that the project area has been identified.

1. The first step is to review existing secondary literature on the types of foods consumed in this area. In addition, rapid appraisal techniques and discussions with key informants can be used to establish a list of foods eaten in the area and coping strategies used by these households during periods of food stress.
2. The next step is to develop a household questionnaire to capture data on a variety of outcome measures of varying degrees of complexity. The measures chosen will need to take into account local conditions and resources (time, money, and people) available for this work as well as the advantages and disadvantages of each method.
3. Data on these outcome indicators are collected.[5] These can be used to provide a characterization of the locality in terms of the nature of the food security problem (is it lack of calories, poor diversity, a problem of seasonal fluctuations in access, unequal access within the household?), the identity of the food-insecure, and the severity of the food insecurity. The methods described above can be used to determine to what extent the simpler measures mimic the more complex indicators.
4. If the association is considered strong, these simpler indicators can be used not only as monitoring measures in their own right, but also as a means of inferring changes in more complex measures.
5. Both simple and more complex outcome indicators can be used to measure impact.

Endnotes

1. The discussion on how to choose indicators can also be applied to process indicators.
2. A fifth method, group rating, is described in Chapter 4.
3. There is no consensus regarding the optimal recall period between 7 and 14 days. In the author's experience, 7 days seems to be the most appropriate. A shorter recall period risks missing foods served infrequently, say on Fridays (in Muslim areas) or Sundays (in Christian areas). A longer recall period can be problematic as difficulties of remembering what was prepared appear to increase. However, other organizations such as the World Bank (in its Living Standard Measurement Surveys) have used the 14-day recall period.
4. A variant of this approach, called a semiquantitative measure of dietary diversity, involves showing respondents pictures or models of different serving sizes of these foods. Respondents indicate whether they consumed the item and in what quantity. From this information, it is possible to obtain a rough estimate of caloric intake. For example, in Honduras, respondents were shown five sizes of tortilla and asked how many of each they had consumed.
5. Chapter 5, on sampling, provides an introduction to this.

4. Rapid Appraisal Techniques for the Assessment, Design, and Evaluation of Food Security Interventions

Gilles Bergeron

Introduction

Project managers in charge of implementing activities that address food-security problems need tools to (1) identify the populations that are food-insecure, (2) design interventions that address the causes of food insecurity, and (3) evaluate the impact of their interventions on the food security status of project beneficiaries. This chapter illustrates how rapid appraisal (RA) techniques can provide useful insights into the research and design of food-security interventions, as well as into the limitations of such interventions. Many factors determine whether RA methods are appropriate in any given case, including the degree of precision required, the characteristics of the population being investigated, and the ability of fieldworkers.

The first section of this chapter presents some general observations about the advantages and disadvantages of RA methods over survey-based methods. The second section presents a set of RA tools that were tested in the field. The tools developed include community mapping, household food security ranking, conceptual mapping of food sources, seasonal food security timelines, and evaluation of an intervention's impact on food security. Each instrument is presented in a similar sequence. First, a brief introduction presents the instrument and its relevance to the study of food security. Second, the tool is described in terms of its specific objectives, format, methods, and products expected. Third, examples from fieldwork experiments are provided to illustrate its use.

RA Methods for Local Needs Assessment, Intervention Design, and Impact Evaluation

Rapid appraisal techniques offer development workers a useful set of research and appraisal tools to quickly obtain information from local populations about their condition and their needs. RA methods also enable local people and outsiders to plan together appropriate interventions and evaluate the impact of development interventions.[1]

RA methods have distinct advantages over survey-based research methods. They generally involve low costs; are highly adaptable to different situations; and tend to facilitate rapport with local communities, which can allow investigators to explore topics not easily studied otherwise or to bring out qualitative aspects that would be missed by surveys. They also favor analysis on the spot with local people, enabling verification of findings and enhancing the local relevance of results. However, RA methods present important disadvantages over more conventional methods, including limited ability to generalize findings, lack of clear validation procedures, and susceptibility to manipulation by informants. In addition, the qualitative focus of RA methods limits researchers' capacity to transform the data, thus constraining the analysis to what is reported by local informants. Besides, the quality of the information collected depends to a high degree on the skills of field personnel.

The general belief that RAs are simple to apply is, in most cases, not true. The selection and training of fieldworkers is much more critical than for conventional enumerators. Finally, because of the use of "participatory-type" methods, RAs tend to raise expectations among the population about program activities. Goals have to be carefully explained from the outset to avoid misconceptions. For all these reasons, the RA approach is viewed in this manual as a complement rather than an alternative to survey-based methods. RA is used to guide, inform the design of, and confirm findings from formal surveys. A combination of formal and RA methods is the best way to ensure the quality of final results.

General Guidelines to the Use of RA Methods

Whenever using RA methods, a number of basic issues must be considered, including:

Training and selection of personnel.[2] As mentioned above, the skill of fieldworkers is critical to the success of RA methods. These skills are quite different from those required by formal surveys. For example, social skills are important: Controlling dominant personalities in group settings while seeking the participation of silent participants—all of this without imposing one's opinions—requires superior communication abilities. Another distinctive attribute is that, unlike survey enumerators who collect data for analysis by outside researchers, RA fieldworkers have to collect, analyze, and validate the data themselves. They are the researchers. Hence they need a sound understanding of the aim of the research so they can, for instance, change the instrumentation used, if need be, without losing sight of the final objectives. The importance of selection and training of field personnel cannot be overstressed. (See the references on training RA fieldworkers.)

Establishing contact. Community life is complex, and care must be taken from the start not to unwittingly alienate groups or individuals by associating too closely with the "wrong" person(s). It is useful to make unannounced visits to a village before the first official visit[3] in order to learn the basic "political language" of that community. This can be done by sending one fieldworker to the village to establish informal contact. Avoiding local authorities is preferable, although not always possible. Free-flowing discussions are initiated with the people encountered, leading to questions such as: Who are the official representatives? How are they perceived? Are there factions or rivalries (political, religious, economic) in the village? Such early knowledge is invaluable when making the first official visit, and helps avoid early missteps.

Then an official visit can be scheduled. In contrast to the first informal visit, this one is well announced and involves local authorities as well as high-ranking officials of the project. This visit is preferably not used for working sessions. Rather, the aim is to explain the project goals and the type of work to be done. Permission is sought from local authorities, dates for workshops are established, and an understanding is established on who will be invited to attend.

Timing of workshop and sequencing of instruments. Project personnel must look for ways to minimize the disruption of people's lives. If possible, the meeting is held in periods or seasons of low activity; otherwise, field personnel must look for a time of day when people are back from their daily activities. Besides showing basic respect, this increases the likelihood that people will actually respond to the invitation and attend the meeting.

The sequencing of instruments during the workshop should normally follow the logical flow proposed in this manual. Some exercises can be undertaken at different moments without affecting the final results—for instance, transects and flow calendars may be done at different times if it is more convenient.

Choice of informants. Initially, all community residents are viewed as potential informants. Some of the exercises—for example mapping and concept definition—can be done without being selective about informants, insofar as they know their community well and are honest in their responses. As the groups most likely to suffer from food insecurity are identified, individuals from these groups soon must play the central role in the discussions. Besides, within identified target groups, subgroups usually need to be considered. Typical subgroups are stratified by gender, livelihood strategy—for example, farmers versus ranchers, age group, and ethnic/caste affiliation. It may be necessary to obtain contributions separately from each group in order to capture all the relevant information. Separating groups may also be necessary if putting them together creates social tensions. The choice of method also must take into account informants' profiles; for example, if the literacy level is low, the method should not require reading skills.

Triangulation. Triangulation refers to the comparison of data between sources to improve the data's validity and reliability. This is particularly critical with RA data (many refer to RAs as "quick and dirty" methods), which are easily manipulated by informants, although group meetings tend to reduce this problem. The important point is that no data should ever go unchecked if they are used for making important decisions. The quality of RA information may be verified in several ways: replicating the exercise with other groups, exploiting alternative sources of information (for example, aerial photos or prior surveys), comparing results against predicted values from mathematical models, "ground truthing" by walking transects, and so on.

Instruments Guide

Concept Definitions

Eliciting local concepts is basic to establishing a common language between fieldworkers and informants. One good time to do this is at the start of each exercise, when the ideas used in this particular workshop are first introduced. The content of each concept is then discussed, so that it is defined in its local, cultural equivalent. Another approach is to hold a special "Concept Definition" workshop where all the notions used in the RA sessions are defined. Whichever method is best depends on moderator preference and on the time available. Appendix 4A provides further discussion.

Approaches proposed to define local concepts go from simple ones, such as brainstorming and pile sorts, to complex ones, such as Delphi methods and cultural consensus modeling. Since all these techniques have the same objective (translating in local terms the concepts used in the RA sessions), the simplest ones should always be used unless compelling reasons require otherwise. Some of the concepts to be defined are described below.

Community. The universe to be mapped has to be clearly defined, so that all households in the village fall within its boundaries and any unit falling outside of it is excluded. Special cases, such as with nomad or pastoral societies that move in and out of the community, have to be discussed and a decision has to be made as to whether or not to include these in the potential target group.

Household. In Latin America, the nuclear family (a man, his wife, and children) is the most common type of household, but in West Africa, extended households (multiple generations/nuclear families living together) are common. The definition of a household may also change depending on whether the focus of the projected activity is production or consumption. If the project goal is production-enhancing, then the targets are the productive units; if the intervention is for food relief, then the targets are the consumption units.

Food security. From the project's point of view, food security is defined as availability and access to food by all at all times.[4] Availability and access, however, are notional constructs that are sometimes difficult for local people to grasp. The following is a useful shorthand for defining these ideas: availability relates to communities; access relates to households. Availability is defined as the capacity of communities to obtain the supplies of food required to

Table 4.1 Realization of the village map

Informants	All villagers/otherwise, selected representatives of the various stakeholder groups in the community.
Where	Large open space. For 3-D maps, preferably outside so the area may be expanded if needed.
Time	Varies with the size of the village and the degree of participation of villagers. On average, three hours should be sufficient to complete the realization.
Objective	Have informants reproduce, at reduced scale, the distinct homes and important living areas of the village. Precision must be sufficient so that all homesteads are clearly identifiable.
Materials	Depends on the type of map and intended durability. No need for fancy materials; instead, use only materials locally available, such as sand, pebbles, sticks, and so on. These are less intimidating than paper and pen for first-time participants. Once finished, the output is copied to large paper sheets or cartons.
Concepts to define	The concepts of community, household, and food security must be defined before starting this exercise. See section on concept definitions in this chapter.
Method	No single method exists for this exercise. Villagers are responsible for its realization and their spontaneous suggestions are encouraged so villagers feel at ease with the instrument and its use. First, a decision is made as to whether a bidimensional or tridimensional map will be done. A tridimensional map takes more time but is more precise, is easier, and is more enjoyable for villagers. On the other hand, time may be short, or the weather may not favor working outside, in which case a bidimensional map should be preferred. Whichever type is used, fieldworkers must ensure that the work proceeds systematically so it has the desired precision. Guidelines to that effect are, first, identify well-known features, such as the central park, the mosque, and so on, and place them on the map. Then, draw the outer limit of the inhabited space in relation to these main features. Next, proceed from the center to the periphery in a concentric fashion. As work proceeds, readjustments to the initial placement of spatial features or to the outer limits of the village are made as required. As households are represented on the map, they are identified by the name of their head. Their characteristics (number of persons in the unit, presence of migrants, number of animals or fields owned, and so on) can also be added at that point.
Products	Two products are generated by this exercise. First (if a tridimensional map was done), the lay model is transcribed on a large sheet of paper, with households properly numbered and identified (if possible, photos of the model should also be taken). All the elements of information present on the map are reported on paper, including names and number of households (note: we assume this requirement is already satisfied if a bidimensional format is used). The second product is a spreadsheet, which organizes the information elicited by the mapping exercise in a matrix format. All items locally associated with food security (for example, fields and animals) that were elicited for each particular household are reported as variables in the matrix. Families are listed as rows, variables as columns. Particular attention goes into coding household identification numbers, especially in cases where extended family units are common (see a model of coding in Table 4.2).
Validation Transects	If high precision is required, an aerial photo may be used.

Source: Compiled by author.

feed everyone that lives there. In a famine situation, for instance, the village's capacity to maintain food supplies collapses. Food becomes unavailable even for people who are wealthy. This is a case where food insecurity is due to low food availability. Access refers to the capacity of households to obtain food. This dimension of food security relates mainly to individual household wealth. For instance, a household that has sufficient land to harvest grain for the full year enjoys greater food security than a household whose land can provide grain for only six months of the year.

Seasons. The Gregorian calendar's month names are not necessarily known to local populations. The length of months or seasons may also vary substantially. The seasons have to be defined before construction of the timeline.

Community Mapping for Census Taking

Community mapping is a versatile tool used to cheaply gather baseline information on a number of indicators—population characteristics, wealth and asset distribution, labor availability, and so on. This manual suggests considering the use of community mapping instead of a formal census (Table 4.1). Besides being quicker, this method may yield better results than a conventional census (but not always—see Christiaensen, Hoddinott, and Bergeron 2001). Another good reason to use this tool is the high level of participation it encourages: villagers usually enjoy mapping, as it is a good way for them to communicate issues that have a spatial dimension. The construction of a map is thus a good starting place for social assessment studies. Note, however, that community mapping is not always the most appropriate tool for census taking—for instance in highly dispersed communities, in areas of low population density, or in situations where the precise targeting of a specific population is of particular concern, a formal census format is preferred.

Example of Community Mapping

Tomba is a community of northern Mali where development agencies are financing the construction of irrigation infrastructure. We visited local authorities, and informed them of our desire to conduct a series of exercises in their village to better understand the local characteristics of food insecurity. The local council accepted a request to map the community, and agreed to invite villagers to participate in this exercise. The time was set for the afternoon of the next day, after they had returned from their daily occupations. A wide-open space, used as a traditional meeting ground, was designated to hold the mapping workshop. We also requested that a selected set of informants meet a few hours before the construction of the community map to conduct a "concept definition" workshop to elicit local definitions of households, wealth, and food security.

The next day, arriving at the meeting place, we were surprised by the level of attendance: all villagers—perhaps more than 200 people—were expecting them. The workshop was obviously seen as a festive occasion, and everyone came in their finest clothes. Field personnel, who spoke the local language, began by explaining the objectives of the exercise to the villagers: reproduce their living space on the ground as exactly as possible inorder to identify household units and the people living in them. The mosque and the central place were laid out first (since these stand in the geographic center of the village), as well as the main paths leading to the central place. Banco (wet clay) was proposed as material, and the staff built a few hypothetical street walls to illustrate the idea.

At the beginning, only two or three men seemed to understand the aim. They proceeded to correct the model. Seeing them work, bystanders quickly joined in and soon all people present, men and women alike, were busy adding their own compound to the map. Controlling the work of so many people soon became impossible, and we were reduced to acting simply as resource persons, answering people's questions about procedural aspects and making sure nothing was left out. As delimitations between compounds were drawn, vigorous discussions were heard all over as to how much of that wall was owned by this compound versus its neighbor, where did this pathway end, and so on. The level of participation, debate, and cross-checking was such that we are confident no major mistakes were made. People clearly counterbalanced one another in making the assessments and little was left unchecked.

Table 4.2 Matrix of household demography, assets, and food security rating: Partial listing from Tomba

Compound number	Domestic unit number	Name of head of domestic unit (HHH)	Gender of HHH	Number of domestic units	Ethnic group	Number of household members	Number of oxen	Owns a plow	Number of cows	Number of goats	Irrigation fields	Non-irrigation fields	Migrant fields	Food security rating[a]
1	1	Abdoulaye Amadou Yatara	1	4	1	10	4	1	1	1	1	1	0	3
1	2	Issa Madiou	1	4	1	8	0	0	0	2	1	1	1	1
1	3	Mamadou Kabara	1	4	1	3	0	0	0	1	1	1	2	2
1	4	Aligui Madiou	1	4	1	4	0	0	0	1	1	1	1	2
2	1	Hamadou Mahamar	1	3	1	3	0	0	0	1	1	1	0	1
2	2	Mahamman Hamadou	1	3	1	6	0	0	0	1	1	1	0	2
2	3	Abdoulaye Hamadou	1	3	1	0	0	0	0	1	1	1	2	2
3	1	Boubacar Madio	1	2	1	10	0	0	0	2	1	1	0	2
3	2	Arsina Madio	1	2	1	2	0	0	0	1	1	1	6	2
4	1	Djougal Iko	1	1	1	4	0	0	0	1	1	1	5	2
5	1	Sidar Traore	1	1	1	5	0	0	0	2	1	1	0	2
6	1	Djoubalo Ahidji	1	1	1	7	0	0	0	2	1	1	1	2
7	1	Aisa Bocar	2	1	1	7	0	0	0	1	1	1	0	1
8	1	Ousmane Kouly	1	1	1	4	0	0	0	1	1	1	0	1
9	1	Ali Oumba	1	2	1	5	0	0	0	1	1	1	0	2
9	2	Hamadou Oumba	1	2	1	2	0	0	0	1	1	1	1	2
10	1	Brema Ousmane	1	1	1	6	0	0	0	1	1	1	1	1
11	1	Hammadou Abdoulaye	1	1	1	8	0	0	0	1	1	1	3	2

Source: Compiled by author from survey data.
Note: Under "gender of head of domestic unit," 1 = male household head and 2 = female household head; under "owns a plow," 1 = yes and 0 = no.
a. See discussion of food-security rating on page 53.

Once the main streets and family compounds had been laid out, people began separating individual homes within compounds by making little clay mounds, each one representing a home. We then asked them to represent their domestic assets, including number of persons present in the home. On each house mound, a number of twigs were then planted to represent how many people lived there—migrant members were represented by a bent twig. Other symbols that represented the household assets were deposited in the yard adjacent to each home. Symbols used included goat feces, to represent the number of goats owned by the home; bean seeds, to represent the number of non-irrigated fields; rice seeds, to represent the number of irrigated fields; and so on.

Once the map was considered complete by informants, field staff proceeded to record the information on a large sheet of paper and the summary matrix was done (see Table 4.2). Particular care was taken when recording family identity numbers, as extended families were common in that village. Compounds were numbered first and domestic units second. Both compounds and domestic units were numbered in ascending sequential order (1, 2, 3, 4. . .), but the numbering of domestic units began anew each time compounds changed. It was also agreed that the first domestic unit named in each compound (which received number 1) would systematically correspond to the family head (Table 4.3). This way of coding was used in order to allow later analysts to associate each domestic unit with the compound it belongs to, a crucial piece of information, given the importance of family networks for livelihood strategies in this region.

Table 4.3 Model used for coding compound and family numbers

Compound number	Family number
001	01
001	02
001	03
001	04
002	01
002	02
003	01
004	01
004	02
004	03

Source: Compiled by author.
Note: 01 indicates family head (HHH).

Food Security Rating

Food security rating is part of a family of field research techniques known as group informant ratings (GIR), which allow fieldworkers to (1) quickly understand how units of interest (households, plots, and so on) are different from each other on a particular aspect (wealth, food security, and so on); and (2) classify them accordingly (Table 4.4). The resulting classification can be used to identify target groups for specific activities. The GIR provide a rapid and low-cost assessment of unit characteristics. In wealth ranking exercises (a popular GIR method), ratings by local informants are further credited with removing the biases of conventional survey methods by bringing intangible elements (such as status, and access to networks of support) to the measurement of wealth and poverty, thus bridging the gap between outsider and local perceptions of poverty.

There are problems with GIR methods, however. The first one is the inability to do cross-community comparisons: Ratings produced are, by definition, contingent on each setting. The GIRs may thus have high internal validity but no external validity whatsoever. Some attempts have been made to overcome this limitation, but no

Table 4.4 Food security rating

Purpose	Classify households in a community according to their level of food security
Informants	Much care has to go into selecting informants. They must be long-standing members of the community, be knowledgeable, and be honest. They should represent a cross-section of the community in terms of age, sex, ethnicity, or other locally relevant distinctions (caste, productive orientation). The number of informants per focus group should be from four to six. Separate groups may be created if members of different social status do not want to stand together in the same exercise, or if women remain silent in the presence of men. Then, however, the ratings produced by each different group have to be reconciled and standardized.
Format	Focus group session
Where	In a calm, private area, inside or outside
Materials	Index cards (as many as there are households in the community plus five for labeling of piles/categories) and markers
Method	Of all the methods proposed in the literature, the "index card" approach is preferred, for it is comprehensive and easy to control. In this method, the name of each household head is written on a separate index card. Once the categories to be used are identified (see "Prior steps" below), a separate pile is created to represent each particular category. Informants talk among themselves and decide which category each household belongs to. If informants are unsure about one household, they put its card aside so the case can be resolved later. Once all households are rated, the moderator takes each pile and reads the names back to the group to give them a chance to review their classification. This may bring additional shuffling. New categories may also need to be created to accommodate intermediary or uncertain cases. If so, all cards have to be read back again to the group, until no more discrepancies are manifested. Once the final categories are made, their attributes are discussed anew, by empirically considering the characteristics of the households falling in this group.
Prior steps	Define the concepts of the community, household, and food security. Define a rating system: informants should be allowed to define their own rating system, so that they feel comfortable with their assessment. Usually, three to five classes are proposed.
Time	About one hour
Products	A listing of all households in the community with their rating in terms of food security categories. A clear definition of what each category of household food security (HFS) refers to.
Validation	Control with attributes of household obtained from mapping. Obtain second opinion from different focus group. Classification and Regression Tree (CART) analysis.

Source: Compiled by author.

convincing alternative has yet been offered. We recommend never using a GIR scale outside the site where it was developed. Second, it must be recognized when GIR is not useful. In communities where everybody is subject to considerable stress, such as is the case with refugee communities, GIR provides spurious or irrelevant details, as differences in wealth or food security become increasingly marginal. Also, the approach is not very useful in large communities where no one can know everybody well. One may divide the larger community into wards or neighborhoods, but then the problem of standardization between subdivisions surfaces (see first point above). Limitations are also noted where populations are highly mobile (such as in pastoral societies), or where households are highly scattered (as in the Amazon). Third, GIRs appear to be very susceptible to error, both systematic and random. Tests of the reliability of ratings suggested that the main sources of error are poor informant selection and poor training of field personnel. This can be remedied by exerting considerable care in the use of the method; however, it has to be clear that it is less straightforward than it initially appeared.

For all these reasons, GIR methods should be used with much caution. They should be used strictly to classify populations within single communities. Careful selection of key informants is required, and careful training of field personnel is an absolute must.

Example of Food Security Rating

A food security rating exercise was conducted in San Marcos, a community of western Honduras where a rural development project is being implemented. The aim of the exercise was to examine how food security varied in the group of farmers targeted by the project. A listing of community members was provided by project managers. We randomly selected various people from that list and visited them, asking who in their opinion were the most reliable and knowledgeable informants in the village. Five persons were repeatedly pointed out by villagers. These five persons—three men and two women—were invited to participate in a focus group session. We explained to them that they would have to create a food security rating of community members. The meeting was scheduled for the next afternoon, and held in the schoolyard.

After informants had arrived at the meeting place, we explained to them what was meant by "food security" and "households" (see discussion above in "Concept Definition" section). They were asked to add whatever they thought should form part of these concepts. Next they were asked two questions: "Does everyone among villagers have equal access to food? (Yes/No)," and "If there are differences, how would you characterize these differences?" After some debate, a two-way classification emerged from these discussions: (1) food-secure, defined as families that never have food security problems; and (2) food-marginal, defined as families that seem to have food security problems every year.

The group was then asked to rate each household on the list in relation to this categorization. The moderator read the names of every household head in turn, asking in each case on which of the two piles this household should be placed. Informants deliberated and then took the card and put it on the appropriate pile. Many cards created difficulties, so they were put aside for later categorization. After the group had gone through all the cards, the moderator asked them to consider again those that created problems. One informant eventually mentioned that it seemed all of them did not fit in either of the extreme categories; rather, they fell in between, not totally food-secure nor totally food-insecure. A third, intermediary category was thus added, which was defined as "families that occasionally have food security problems but not every year." The moderator added a new corresponding pile. He then read back the names that had been put on the two first piles (Food-Secure and Food-Marginal) and asked if they still agreed on this rating. Many of the households from these piles were then reclassified to the intermediary category.

Once the review was completed, the moderator asked informants to consider again each class and the households in it, and asked, "What makes you think these households belong to this class?" Responses to that question improved understanding of food security differences in the community, and provided a point of entry for later project design. Mentioned characteristics were as follows:

Food-secure group

- They work at a large scale on their own lands.
- They have good ideas.
- They work hard.
- They save their money.
- They have the best lands.
- They have public responsibilities.
- They have cattle.

Food-insecure group

- They do not have much land.
- They have to work for wages occasionally.
- Their families are large, and the little they produce is consumed right away.
- They sell their product before it is harvested.

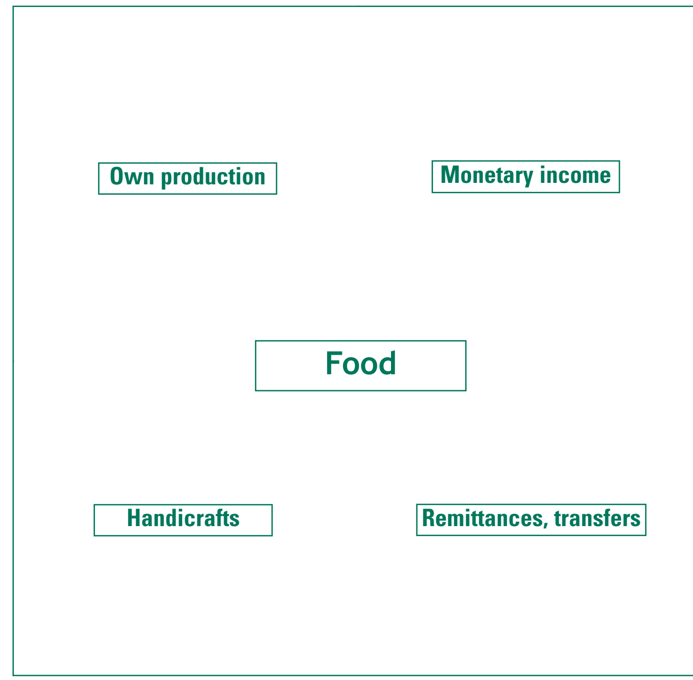
Food-marginal group

- They always have to work for wages.
- They have no money, low revenues.
- They do not make decisions; they do not have a view of the future.
- They are lazy or sick people.
- They do not have a sense of responsibility.
- They must buy all their food.
- They do not have land, or their land is insufficient.

Conceptual Map of Sources of (and Threats to) Food Security

Conceptual mapping is a relatively new technique in the participatory rural appraisal (PRA) tool set, used to specify which factors contribute to a particular outcome. It can be viewed as the qualitative version of a functional equation in which the outcome (dependent variable) is determined by a set of factors (independent variables) that can be objectively specified and ranked in terms of their respective contribution to total explained variance (Table 4.5 and supporting Figure 4.1a–c).

Table 4.1a " Zoning" of the conceptual map into quadrants



Documented experience in the use of this technique is scarce. Our field trials suggest that, although theoretically promising, obtaining good empirical results is a challenging task. We noted two main difficulties. First, the map is complex and requires a very skilled moderator. Second, verification is problematic: Supporting evidence is difficult to obtain and requires a better knowledge of the community

Table 4.5 Conceptual map of food sources and threats to food security

Purpose	(1) To elicit the most important pathways by which households obtain their main staple food in that community, (2) identify the most important threats to these food acquisition strategies, and (3) assign priorities to these threats.
Informants	Optimal size of group is from 8 to 12 participants. Informants must be selected to represent the distinct farming strategies found in that community. A balanced gender representation is also required.
Format	Focus group session held in a quiet, private area.
Materials	Materials include a large sheet of paper and markers of distinct colors.
Methods	This exercise is easier when limited to main staple foods (for example, maize and beans). ***(a) General aspects.*** The moderator explains to the participants that he/she wants to know the sources of their staples in this community. A simple example (for example, "growing it") is usually sufficient for participants to understand what is expected from them. Informants will mention that they get staples from their own production, donations, purchases, and so on. Always remind informants to refer only to actual, nonhypothetical sources of food. Also, a minimal number of families—for example, at least 25 percent of households—should use this strategy before it gets recorded on the map. ***(b) Mapping food sources and their pathways.*** The moderator "holds the pen" during the whole session, so the product remains organized as it fills up. The moderator mentally divides the map in "zones" to keep sources separate from one another. An example of "conceptual map zoning" is presented in Figure 4.1a. Once the main sources of staples are listed, each source is considered individually. The main prior conditions to this source are elicited. For example, a prior condition to have "food from own production" is that there be a harvest. To have a harvest, the farmer must have land and buy inputs. Both of these require capital, which may come from savings or loans; and so on. Each of the steps in this sequence corresponds to a node; the full sequence of nodes associated with a particular source is called a pathway. The pathway and its nodes are reported on the map as in Figure 4.1b. ***(c) Ranking food sources by order of importance.*** Conceptual maps generally turn out to be very similar from one village to another. What makes them different is the relative importance of each pathway in the livelihood strategies of the villagers. Once all pathways have been identified, a subjective weighting is made between them by drawing arrows of various sizes indicating their relative importance in the community. The size of each figuratively corresponds to that vector's effect. ***(d) Identify threats to each food source.*** The moderator next asks informants to identify the main threats that exist along each pathway. The link between each node is examined, and elements that may threaten this link are elicited and written on the map, using a marker of different color. Here again, it is important that the threats identified correspond to those that exist in this village, and not merely theoretical ones. Since threats are usually different between sites, the map will also differ between villages at this level (see Figure 4.1c). ***(e) Prioritize threats to address first.*** The final step is to rank threats by order of importance. Pairwise ranking is adequate for that purpose.a To keep this manageable, a maximum of five threats per pathway is suggested. If three pathways are identified, that makes a total of 15 threats to rank.
Prior steps	Identify main staples. Recruit informants.
Time	Approximately two hours.
Products	Products include (1) a specification of main staple sources in the community and their relative importance, (2) an identification of the main threats to these pathways, and (3) a list of threats in order of priority.
Validation	The only rapid way of validating the results is to repeat the exercise with another group and triangulate findings. A household-based survey of food consumption may provide information about sources of food, but not about pathways or threats. A prolonged stay in the community (six to seven days) is needed to verify the conclusions.

Source: Compiled by author.

a. Pairwise ranking is a common RA technique in which every choice is iteratively compared with every other choice by asking which of the two is most important. In this way, all choices get ordered in terms of their relative importance one to the other.

Table 4.1b Nodes and pathways in conceptual map

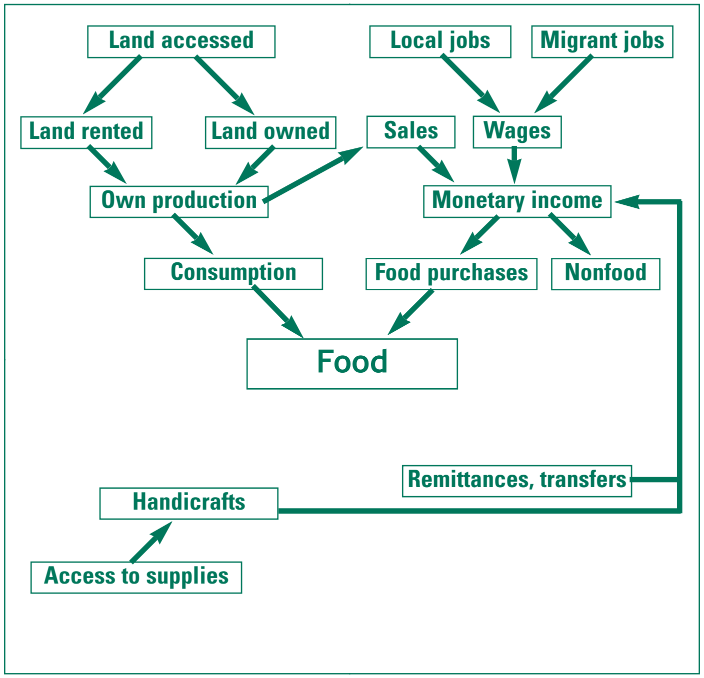
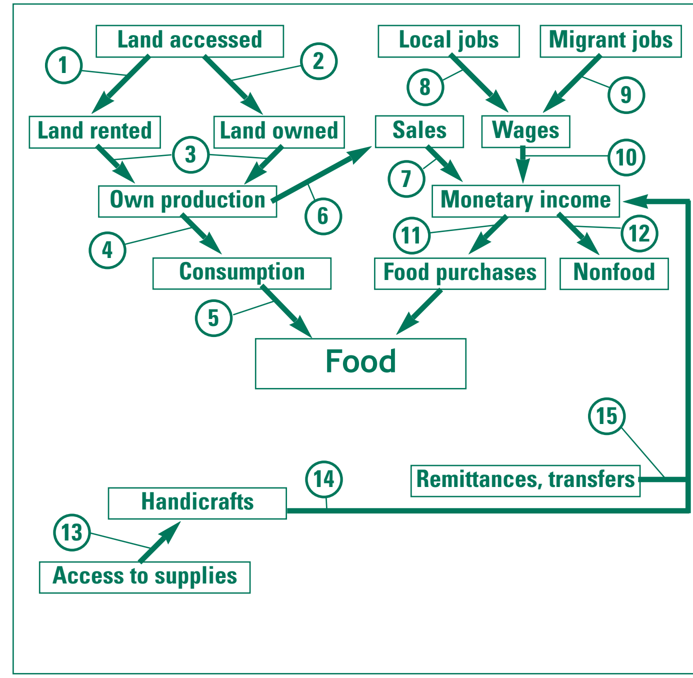


Table 4.1c Threats to food pathways



1. Availability of land for rent
2. Price of land
3. Rain, pests, access to capital
4. Storage losses
5. Capacity of household members (HHM) to use food
6. Output prices
7. Total output
8. Demand for wage labor
9. Availability to travel
10. Wage levels
11. Prices of food
12. Prices of nonfood items
13. Arrangements to access materials
14. Handicraft output prices
15. Presence of migrant member or access to income transfer

Source: Compiled by author.

than the little time spent doing a rapid appraisal can actually generate. Yet, this exercise can be very useful for assessing the sources of (and threats to) food supply. For this reason, project managers should be aware of its potential when exploring options for food security interventions. Basic guidelines about its use are provided below. It must be emphasized, however, that it should be used only if qualified personnel and time are available.

Example of Conceptual Map

Main Food Pathways and their Prior Conditions

Santa Teresa is a mountain community of western Honduras. Staple foods are maize and beans. Villagers obtain these staples either by producing them or through purchases. No food donation programs are active in this community, and few households mention receiving transfers. Staples include maize and beans, grown primarily for subsistence with small quantities occasionally sold locally for cash. Wheat was once important, but less of this crop is grown every year due to genetic erosion, and the small amounts produced are grown only for sale. The prior condition to production is access to land, labor, and inputs. Land in this village is either owned or rented by the producers. Labor depends on the family demographic cycle. Inputs are generally bought, since organic fertilizers are little used locally. The working capital for production comes from credit, savings, the sale of produce, or from wage work.

Food purchases depend on income generated from two distinct sources: The sale of one's own production and wage work on other people's land. The conditions that determine sales are the same as those determining production. Thus, land access is the key to how much cash is derived from production. Wage work refers mainly to temporary migration during the coffee harvest season.

Threats to Food Acquisition

Pathway 1–own production. The local production of basic grains is determined by many factors. Farmers say external inputs are crucial to their production of food. Most of the money to buy these inputs comes from loans; but to obtain a loan one has to own land, be a member of a producer organization, and be free of debt. In Santa Teresa, about half of the people own some land. They recently formed a producer association, enabling them to access credit. For them, the conditions to access credit are met—unless they have bad loans. For those with no land, however, the situation is more difficult. They may rent land, but rented land cannot be used as collateral and does not give access to credit. Besides, land rental is insecure because of the legal stipulation that a farmer who has worked a plot for more than three years can claim ownership of that plot. Fearing loss of their ownership rights, landowners prefer not to rent to the same person from one year to another. Landless producers thus constantly have to seek new land to sow their basic grains. This lowers the incentive to land investment, and rented land is typically more degraded and of poorer quality, making it (and the family that uses it) more vulnerable to production shortfalls. There are few ways out of this situation, as the land market is tight in this area, and buying land is expensive.

Assuming land and capital are secured, the next problem confronted by producers is the price of inputs, which is always increasing. This complaint is certainly legitimate in the case of basic grain producers. Other sources confirm that the cost/benefit ratio in basic grain production has gone down in Honduras by up to 40 percent in the last two decades (compared with an increase of more than 200 percent in nontraditional commercial crops). This is bound to have severe effects on a community like Santa Teresa, where people rely to a high extent on their own production to ensure their supply of basic foods.

Going down the pathway, and assuming fertilizers are obtained, farmers still have to face the hazards of erratic rainfalls, pest outbursts, postharvest losses, and so on. Irrigation systems could remedy rain shortages, but water sources are distant and would have to be pumped, requiring a major infrastructure investment and high operational costs. Pest incidence is relatively low in this community, yet pesticides are needed at times, which again requires capital. Storage losses, largely from rot and rodents, are reported to affect up to 15 percent of stored grains.

Pathway 2–purchase of foods. The capacity to buy food is related to the wealth of a household, which is a function of the amount of land owned, sales from one's own production, access to savings, and/or earnings from wage work. The threats associated with production were already described. To these, one must now add the problem of output prices, which fluctuate quite dramatically on a seasonal basis. With respect to wage work, the most important source of employment is provided by coffee harvests. However, this source of income is premised upon the availability of household members for periods of out-migration and the effective demand for labor in the coffee sector, which is a function of world coffee prices and climate. Coffee harvests occur only in a short, seasonal fashion, but the incomes provided are secure and stable. Yet farmers resent this obligation to migrate, and they would rather stay at home if they could. Also, they complain that salaries are low (although other sources report that coffee wages have improved over the last few years). A few alternate sources of employment exist locally, but they are occasional and cannot serve as a main source of income. They also pay less.

Finally, producers mentioned that the purchase of food is affected by problems of local availability (nonexistence locally) and access (high prices). Prices, they say, are particularly subject to manipulation by intermediaries.

Table 4.6 Matrix of threats to food acquisition, with possible actions and their likelihood

Problem	Possible action	Likelihood of action
Inadequate tenure laws	Change land tenure law	Unlikely: Tenure laws are a national policy.
High land prices	Change land market Land reform	Unlikely: The market is already quite open. Local land reform would provide no relief, as landowners in this community are smallholders.
Production hazards	Stabilize yields via technical improvement	Can be done. Technologies can be adapted to improve maize/beans/climate/pest tolerance.
Poor access to capital	Offer credit without need for collateral	Can be done, but requires organization. Alternative credit guarantees—for instance, group lending—must be explored.
Storage losses	Provide silos	Can be done: Simple, cheap technologies exist.
Poor or unstable output prices	Diversify in high-value crops to deflect poor prices of basic grains	Diversification into commercial-output-prices crops might be envisioned, although this needs to be paired with irrigation and roads for market access.
Poor labor market	Stabilize labor market	Unlikely: Local outlets are saturated and there is no control over demand for labor in coffee.
Poor wages	Improve wage levels	Unlikely: Wage levels are determined nationally.
High food prices	Remove middlemen via consumer co-op Favor production of vegetables in home gardens Improve transport	Possible, but difficult. Consumer co-op requires much organization and training. Can be done. Additionally, favors involvement of women and children in food production and offers alternative source of income and sales Possible, but costly. Could be paired with consumer co-op.

Source: Compiled by author from survey data.

Analysis and ranking of threats

The threats identified above were listed for further discussion. A matrix (Table 4.6) was drawn to discuss the possible action, and whether any of these actions were in the project's and the community's manageable interest.

A pairwise ranking was made to prioritize issues to be addressed by development agencies. The following were listed in order of preference:

1. Offer creative solutions that would provide credit funds without need for collateral.
2. Make technical improvements for yield stabilization in basic grains.
3. Construct storage silos.
4. Diversify production towards higher value crops.
5. Favor production of vegetables in home gardens.
6. Create a consumer coop to remove middlemen.

Seasonal Food Security Timelines

Diagrams such as pie charts, bar graphs, and timelines are very popular among rapid appraisal workers seeking a chronological representation of processes. Considerable documentation is available on the various types of chronological instruments that have been developed and their uses (see References). The timeline is a particular version of these that models time-bound processes in a linear fashion (Table 4.7). Timelines are very flexible: One can find applications all the way from history manuals, where they are used to describe long historical sequences, to software planning tools, where they are used to describe sequential flows of activities in a project. In this guide, the technique is used to better understand the sequence of events leading to food insecurity. To do so, multiple timelines are superimposed to illustrate the connections between production and consumption flows, and cycles in asset availability and demand for cash. The data thus provided can be used at distinct phases of project design: in initial needs assessment ("When is the hungry season?" "What food runs out first?"), project design ("What combination of early/late maturation breeds could reduce the length of the hungry season?" "When is labor available to realize projects?"), and evaluation ("How do calendars compare between the beginning and the end of the project?").

Example of Timelines

Data from the community of Santa Teresa in Honduras illustrates the use of timelines (Table 4.8). As already noted, mountain wheat is produced in this community in addition to the usual Honduran staples of maize and beans.

Harvests. Food harvests go from August to January, but they are divided in two distinct subperiods: August and September, and November to January. The little wheat that is still harvested comes mainly in September, although a few households also obtain small amounts of wheat in August. Maize harvest begins in November, increasing gradually until the peak month of January. Small amounts of early maize (elote) may be harvested also in September and October. Most beans are harvested in December, with small amounts coming up in November.

Monetary revenues. Monetary incomes come mainly in the last two months of the year (November and December) and in the first three months of the year (January to March). Cash comes either from the sale of one's own production (wheat in a few cases, which is sold in September, and maize, in most cases, sold between December and March, with sales culminating in the latter month), or from wage

Table 4.7 Seasonal food security timelines

Purpose	Describe yearly cycles of food production, food consumption, cash, and labor use.
Informants	Two different groups are consulted: A set of community informants chosen from the whole village to develop a typical community timeline; and a set of households viewed as most food-insecure, to develop timelines for food-deprived units.
Format	Focus group sessions held in a sheltered, private area.
Materials	A predesigned matrix (months as columns and flows as rows). Six groups of seasonal flows are considered: harvests, income, expenditures, labor, food, and cash. Each is further subdivided into single categories: *Harvests:* Consider individually the three main crops grown locally, at least one of which is a staple (the other two may be cash crops or staples). A rating of their relative importance in terms of the amount of labor they require is also provided. *Cash income:* Distinguish between income sources from agricultural sales, wage work, and sales of handicrafts. Their relative importance is also rated. *Cash expenditures:* Distinguish between production and consumption expenditures. Include only recurrent, important ones. For each, consider the total amount of cash needed, for example, for production expenditures. Informants must add costs of inputs, hired labor, animal medicine, and so on. For consumption, they must consider the need to buy food when supplies from their own production are over, plus school materials. They then add all of these in deciding when more money is needed. The calendar reports on the total. *Labor:* Includes mainly timing of female labor. Could be divided between labor in own farm versus labor for wage. *Food and cash:* Describes periods when food and cash are scarce. Markets are chosen to be representative of the cycle being described (coffee, maize seeds, bean seeds, and so on.)
Method	The calendar is laid on the floor, and participants are invited to stand around it. The purpose of the exercise is explained, and the moderator indicates how to use markers. The exercise begins with the harvest of the most important subsistence crop in that community (first row). Say it is maize. The moderator asks participants, "In which month do you mostly harvest your maize?" One of the participants is asked to put five maize grains in the cell corresponding to the designated month. The moderator next asks whether harvests of this crop are obtained in other months. Another number of maize grains are deposited in the corresponding cell. It is explained to participants that the number of grains corresponds to the relative amount obtained in each month, so that months with greatest harvests have the largest numbers (five) and those with smaller harvests have the smallest number (one). Intermediary months may receive from two to four grain markers. Months without harvests are left blank. Each timeline is revised in a similar fashion, that is, the month of greatest importance receives the largest number of markers, with the exception of "months where food and cash are scarce." These are inversely classified to indicate periods of greatest scarcity (months of greatest scarcity get more markers). After the community workshop, the exercise is repeated with the three most food-vulnerable families (selected from Food Security Rating results). In this case, however, the timeline is made specific to those households' situations. The objective here is to assess how these households compare with the rest of the village.
Prior steps	Identify the main crops and income-earning activities in the community. Identify informants from the food-insecure group. Describe seasons in local words.
Time	Approximately one hour per group.
Products	Once finished, project staff transcribes the result on a separate sheet, coding the size of mounds from 1 (smallest) to 5 (greatest). Pictures are taken of the final calendar if possible. Relevant details that do not get reported on the timeline are collected by the relator, to be reported later at the time of write-up.

Source: Compiled by author.

work during the coffee harvest season, beginning in November and culminating in December and January. Some additional wage earnings are obtained in February, mainly obtained from working in coffee harvests, which implies seasonal migration. No other sources of cash are reported; trade or handicrafts are not mentioned.

Women's labor. Women do not work in other people's fields. They work only in their family's plots. Their involvement in agriculture occurs in two periods: land preparation for maize in June, and maize harvests in December and January.

Expenditures. Most production expenditures occur at the time of land preparation, before the sowing of maize (May–June) and shortly after fertilizer or weed killers are needed (August–September). Consumption expenditures concentrate in the months from June to August, with a culmination in the latter month, when foodstocks are exhausted and school equipment has to be bought.

Food reserves and monetary savings. Food reserves usually last until June. From that moment on and until September, when a few early maize cobs can be harvested, people depend almost entirely on their monetary savings to buy food. Monetary reserves reach their lowest point between the months of June and August, but the period of scarcity may begin as early as April or May. The early maize harvests in September provide some relief at that point, if the season is favorable.

Summary of the timeline. In summary, the timeline indicates that the supply of food is at its highest between the months of November and January. Starting between April and June, we note a

Table 4.8 Development projects: Multiple timelines form (example from Honduras fieldwork)

Community: **Group: (Mixed, Males, Females, Individual)** **Date:**

Category	R*	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Harvest of (main crop)													
Maize harvest													
Bean harvest													
Income from production													
Income from wage work													
Income from other work													
Women's work in own farm													
Women's work outside													
Production expenditures													
Consumption expenditures													
Low reserves of food													
Low reserves of cash													

Source: Compiled by author.
Note: R refers to relative importance within each category.*

progressive decline of food and monetary resources, which culminates in August when severe scarcity is mentioned. The small harvests of maize recorded in September alleviate this situation; from that point on, food access and food availability improve progressively until the cycle begins again.

This sequence indicates a high level of dependence on the maize harvests in September and afterwards. The total maize harvest can be assessed by the end of January, and dispositions could be taken to alleviate future food shortages based on an assessment of total harvests at that date. Another indicator of future harvest performance is the quality of the rainy season. Late or poor rainfall (which can be assessed by July) can create a difficult situation for the coming September and October, translating into a serious problem of food access and/or availability. A combination of these two situations can be disastrous. A monitoring of the situation at these two critical points would be useful to forestall severe food security problems.

Production expenditures occur mainly in the rainy season (May to August). Credit funds must be available in these months if they are to affect the current growing season.

Comparison with poorer households. The same exercise was carried out with three households identified as food-insecure by the Food Security Rating exercise. Similar situations were reported by those three households. Differences with the general village situation were particularly evident along the following lines:

- In all cases, fewer months of harvests were reported, no matter the crop. In two cases, no maize of segunda (second crop cycle) was obtained, and none grew wheat.
- Income from sales of their own production came in fewer months, if at all, and meant little. By contrast, income from outside sources was important. Wage work in a traditional tile factory was cited by one as a main source of income, day labor in coffee farms by another.
- The time spent by women working outside the home was much greater in two of three cases. In both cases, women worked for wages, not on their own farm. The third case corresponded to an elderly couple, who reported no sources of income at all (they subsisted on transfer income from charities).
- The period for production expenditures was much shorter in all cases.
- The months of scarcity were approximately the same, but extended for longer periods.

It was clear, from conversations with these households, that their main problem was lack of access to land, but also to labor and other productive resources. None of them owned land, and two rented small plots on a yearly basis—thus the little amounts of produce reported, either for consumption or sales. This lesser emphasis on agriculture also explains the different timing and direction of expenditures—little went to production, most went to subsistence. Women's work is certainly of concern, as this may lower their ability to care for younger children, without apparently bringing compensation in the form of sufficient income.

The lesson from the timeline is that quite different strategies might need to be envisioned if the project is truly interested in dealing with food security issues. Alternatives to agricultural production—for example, value-added transformation of locally produced goods—may do more for those particular families than agriculture-oriented interventions. The best strategy would be to combine both.

Monitoring and Evaluation Workshop

The last exercise aims at monitoring and evaluating the impact of the project on local food security (Table 4.9). It is conducted at least one year after the beginning of the project, so the activities have time to manifest some impact. It may be done on a yearly basis thereafter, to assess whether the project is on course and enable changes if needed (monitoring function). It may also be realized at the end of the activities, to draw lessons and guide the design of future activities

Table 4.9 Monitoring and evaluation of impact

Purpose	Monitor the progress of activities with respect to stated goals, and evaluate the overall impact of activities at completion to inform, orient, and improve design.
Informants	Beneficiaries of project activities.
Format	Focus group session including 8–10 informants, held in a quiet, private area.
Materials	Large chart prepared in advance, listing activities in rows, and whether they had an impact on income, food access, and food availability in columns. The last column is left for explaining reasons for impact or lack thereof. See Appendix 4B for an example.
Method	• List the activities undertaken by the project in that community (list only activities that have been implemented, and which had time to have an impact; for instance, the impact of a tree nursery on community life will not be felt before some years, so this activity is not evaluated). This list may be obtained from project officers working in the community. It is later validated with local informants in the village to ensure that the activities noted in project paperwork indeed correspond to those deployed in the community, and that no important activity is omitted (or added). • Considering each activity in turn, ask villagers whether this activity had the effect of increasing income, food access, or food availability in the community. A good definition has to be provided for each of these notions. Access refers to the food obtained at the household level. Availability refers to the food found at the village level. Income refers to cash earnings associated with the activity (see definition above). • Informants are asked about the reasons for the success (or failure) of the activity. For instance, if the activity is technical extension in maize production and villagers report lack of impact on food access in the first year, this may be due to poor implementation of the activity, but it may also be due to poor rainfall or to a pest outbreak. Likewise, the failure of a credit program may be due to a late delivery of funds, but also to the unavailability of inputs locally. The actions listed in the project paperwork can be consulted to augment this characterization (that is, each activity is supported by specific actions). In case the activity is not successful, we may ask whether the actions were indeed taken, and the failure to do so may explain why the activity did not have any impact. • The activities considered most successful (in terms of villagers' priorities) are listed, followed by the less successful ones, and so on, until all the activities have been listed and ranked in relation to one another. • This exercise is also undertaken with the technical staff in charge of the program. Comparing assessments between project managers and beneficiaries validates the findings and provides a more complete and balanced evaluation of the activities.
Prior steps	Identify the main crops and income earning activities in the community. Identify informants from the food-insecure group. Describe seasons in local words.
Time	Approximately one hour per group.
Products	Once finished, project staff transcribe the result on a separate sheet, coding the size of mounds from 1 (smallest) to 5 (greatest). Pictures are taken of the final calendar if possible. Relevant details that do not get reported on the timeline are collected by the relator, to be reported later at the time of write up.
Validation	Repeat the exercise with another set of informants and compare results. Plausibility should also be corroborated with external data.

Source: Compiled by author from survey data.

(evaluation function). Note that this exercise does not aim to replace the monitoring and evaluation procedures based on the collection and analysis of quantitative data by the project. Rather, the aim is to ensure that the voice of local people is heard and that their opinions on the activities and suggestions for improvements are taken into account.

Here again the we found no documented experience in the literature on this topic, but experimental trials in certain project sites proved satisfactory. It is estimated that there are two crucial requirements for a successful completion of this exercise. First, only the direct impact of activities is evaluated. Second, the outcome variables are the components of food security (that is, food access and food availability). Income is also considered an outcome variable, as many activities directly target income, and income indirectly affects access or availability. These three dimensions are defined to the participants as follows:

1. Increased income refers to additional sales resulting from increased production.
2. Increased food access refers to the greater presence of food at the household level, and results when more food grains are produced as a result of project activity.
3. Increased food availability refers to the greater presence of food at the village level, and obtains when the activity results in additional food being sold in the village, thus augmenting the amount of food available in the village as a whole.

For instance, technical assistance in coffee production may result in increased income, but not in increased food access nor food availability, as coffee is not eaten. Only through the increased income generated by coffee sales may food access be improved—but it may not have this result, since the increased income may not be spent on food. Thus, it is important to identify only direct impacts. As another example, if the project improves bean production, and this increased production is both sold and consumed, then the assistance will have an impact on incomes, on access, and (if beans are sold locally) on availability.

Example 1: Using the Impact Evaluation Instrument

The example of Santa Teresa illustrates the use of the Impact Evaluation tool. The community had been visited the previous year by an NGO. This NGO had identified the following objectives for its activities in that village: increase maize yields (no target specified), increase bean yields, improve handling of minor species, train villagers in environmental protection of water sources, train villagers in proper use of credit, and implement a credit program.

Increase maize yields. Villagers say this goal was reached. Their maize yields were higher this year than in previous ones, although the precise improvement was not known. This yield increase had positive effects on food access, mainly via the augmentation of subsistence production. It had very little effect on either income or food availability, however, since only a few households sold maize.

The increase in yield was due to (1) a favorable rainfall in that season; (2) the training farmers had received from the NGO in improved seed selection, better agronomic practices, and proper use of fertilizer; and (3) the availability of credit for purchasing inputs.

Increase bean yields. Bean yields were reportedly higher this year than in previous ones. This goal was reached, although again the exact improvement is not known. The increase in bean yields had positive effects on income (in Santa Teresa, beans are as much a cash crop as a staple), on food access (households' production of this staple went up), and on food availability (more of the production was sold locally).

The reasons for improved yields were similar to maize: improved agronomic practices, and better fertilization and pest control practices. Farmers also received improved genetic materials through the NGO. Favorable rains also helped production. Farmers also received credit, which allowed them to buy the inputs they had been taught to use by the NGO's agronomist.

Improve handling of minor species. No activities were developed around this objective, so it had no effect on any of the three outcomes. Villagers said they did not know why the NGO had left aside this part of the work plan. When consulted, the NGO staff said their contract with their funding agency had come to an end, and no resources were available to develop this aspect.

Train villagers in environmental protection of water sources. The same situation as for training in minor species was reported on this activity. No training took place, and plans for reforesting riverbanks were left undone. Here again, the NGO blamed this on a miscommunication with their funding agency representative.

Train villagers in proper use of credit, and implement a credit program. Credit principles were taught, and villagers said it was very useful. Part of the training consisted of creating a producer association responsible for channeling and administering the individual loans. The creation of this association had secondary benefits, such as providing a conduit to farmers' requests for technical assistance and providing a focal point for the realization of public goods activities like road repairs, soil conservation structures, and so on. Thus, although this training had no direct effect on incomes, food access, or food availability, it was undeniably beneficial to the long-term well-being and food security of Santa Teresa's inhabitants, as it incited better community organization.

Credit was obtained in the last production season. The effects on outcome indicators were indirect, but villagers say it had a critical influence on final yields.

Example 2: Using the Impact Evaluation Instrument

The impact evaluation instrument can also be used by project managers to evaluate how well they are doing globally, how well particular classes of activities serve the objectives of improving food security, and how well particular NGOs are doing in implementing their contract. To illustrate this, results were compiled from 10 communities of western Honduras where a number of NGOs implement development activities. A total of 16 types of activities were carried out across all communities—note, however, that none of the communities hosted more than 8 activities in total. Table 4C.1 in Appendix C reports on the results, breaking down by village (columns) and activity type (rows), each type being, in turn, divided by its impact on income (Y), food access (AA), and food availability (DA). An additional line specifies the NGO in charge of this particular community. Examination of the table offers the following insights.

- The overall rate of success was 33 percent.
- The three most successful types of intervention for improving income were agronomic training in coffee production, credit programs, and agronomic training in bean production.
- The three most successful types of intervention for improving food access were agronomic training in maize production, training in care of minor species, and agronomic training in bean production and diversification of production.

- The three most successful types of intervention for improving food availability were diversification of production, training in care of minor species, and agronomic training in bean production.

This information suggests that the overall rate of success is rather low. This assessment is tempered by many factors, however, as revealed by detailed consideration of the data. First, it seems that agricultural production-oriented interventions usually work well. Other types of interventions by contrast—improving commercialization, inciting alternative income-generating activities, protecting the environment—do poorly. Project managers should thus consider whether to emphasize these types of activities in the future, or (given their poor rate of success) abandon them altogether. In making this decision, due consideration should be given to the guidelines emitted earlier to direct NGO work, and whether the conceptual tools were available to them to develop this type of activity.

Other elements may explain the poor overall rate of success. First, many activities have been implemented only recently, and have not had time to manifest their impact yet. Thus the same assessment should be made again at a later date—say, in one year—to see if the patterns documented here hold over time. Second, and unlike the example in Santa Teresa, many communities suffered from adverse climatic conditions in the last production year, and this may have thwarted any beneficial influence from the interventions.

Appendix 4A: Methods for Local Concept Definition

In this appendix, we review a few of the most important techniques used to identify and define local concepts. Three techniques are examined: cultural domain identification (or free listing and pile sort), Delphi analysis, and cultural consensus modeling.

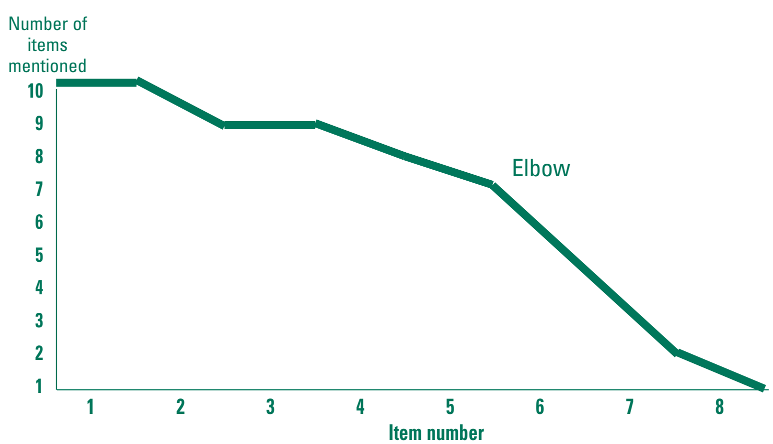
Cultural Domain Identification

Practically speaking, to define a cultural or cognitive domain is to make a list of its elements. Such a definition is needed when one has a general idea of the domain, but does not know exactly which items belong to it in the particular society under study. To determine this, anthropologists commonly use free listing techniques (akin to brainstorming sessions), in which a set of respondents is requested to name all items matching a given description.[5] For example, if interested in the domain of "food vulnerability," one asks each informant to individually identify all the elements he or she associates with that term (it may be working for wages, or lacking land, but also may refer to traits that are specific to that culture, for instance, being in a casted group, or not having a camel, and so on). Once the brainstorming has elicited all the attributes associated with the term of interest, the list is further processed using particular techniques, such as "pile sorts" and "ratings." They consist in simply counting up the number of times each item is mentioned, and sorting the list in order of decreasing frequency. A well-understood concept (for example, one that informants easily associate with their daily lives) will typically have a core set of items that are mentioned by many respondents, plus a large number of items that are mentioned by few or just one person. It is assumed that the core set of items reflects the existence of a shared cultural norm regarding that concept, while the additional items represent the idiosyncratic views of individuals (Borgatti 1993). The shared cultural norm is what is of interest.

The first step in distinguishing the "shared" from the "idiosyncratic" is a distribution of the frequency with which brainstorming items are mentioned. If represented in a scree plot, the cutoff point between shared and idiosyncratic items should be indicated by a drop (or "elbow") in the plot. In Figure 4A.1, for instance, items 1 and 2 are mentioned 10 times each, and others with declining frequency. The elbow method suggests a natural cutoff point between item 6 (mentioned 7 times) and item 7 (mentioned twice). The concept here is thus formed by the six first items. If no clear elbow shows, then one can pick the top *n* items, or items that are listed by more than *x* respondents, as the cultural definition of the domain.

Whatever the rule used to eliminate noncore items, one should always ask why some respondents did not mention items that were commonly mentioned, or that were theoretically expected to be associated with the domain. In many cases, the reason why an informant does not mention a particular criterion may not be that it is irrelevant, but that it did not occur to him or her at the time of the questioning. Such "informant blanking" can be rectified through more discussion. If the variation in frequencies is due to real individual differences in opinion, however, then more steps are needed. The researcher should first make sure that the concept is clear to the informants. A concept like "food security," for instance, may be diffused, and need to be reformulated before consensus is reached on its local meaning. It may also be that the concept per se is unfamiliar to local people. An example of this situation arose in Guatemala when indigenous people were asked about their natural resources conservation methods. The informants did not understand the question because conservation exists as an intrinsic part of the farming system, not as a set of activities independent of it. If it is concluded, as in that example, that the lack of concordance on a concept is due to the absence of a precise cognitive referent, then the researcher should resort to one of the other strategies listed below that rely more on "specialists" (people who understand this problem because of their particular situation or knowledge).

Figure 4A.1 Scree plot of core items



Source: Compiled by author from survey data.

Delphi Method

The so-called "Delphi method" is an iterative definition process designed to achieve consensus among a group of persons considered experts on a particular topic as to the criteria used in evaluating this topic. This is especially useful in situations where no standard criteria yet exist for doing this evaluation. The method is well documented and has been used in a wide number of applications.

The procedure consists of the following steps. Begin by identifying a set of "experts," or individuals who have a vested interest in the issue. Then each is asked a few questions, following a standard format. For instance, assuming that the two areas of interest are criteria for evaluating food security, and criteria for evaluating causes for loss of food security, these questions could take the following form:

Question 1. Assume you are in the middle of the dry season. Please list the five most important criteria you would use in assessing your food security situation on that day from your own point of view (that is, as a cattle rancher or as a coffee grower). Once you have made your list, please rank each of these criteria from one to five, with five being the most important factor. Give reasons for the importance given to each factor. Also, give an an opinion as to how each could be measured.

Question 2. What are, in your opinion, the five most important reasons for deterioration in food security? Once you have made your list, please rank each of these criteria from one to five, with five being the most important factor. Give reasons for the importance given to each factor. Also, give opinion as to how each could be measured.

The next step is to reduce the quantity of information provided to a manageable number of criteria. This is necessary because of the large number of responses that may be elicited. A large number may be useful in terms of domain mapping, but it is impractical in terms of establishing streamlined evaluation criteria. To reduce the impact of too many responses (and also to reduce the impact of informant blanking), a second round of questioning is done, using the same cues, but asking respondents to select among the list of criteria elicited in the first round. Respondents are also informed that they do not have to list the same ones as before; rather they should consider whether any of the criteria mentioned by others would be a better criterion than any of those they originally proposed. This procedure has been demonstrated to drastically cut the number of criteria mentioned. Finally, the most important criteria are isolated using individual criterion scores, ranking them from most important to least important, using a five-point Likert scale. The final list of valuation criteria may be finally reduced to the five or 10 most important ones, according to this last ranking exercise.

Cultural Consensus Modeling

Cultural consensus modeling describes and measures the amount and distribution of cultural knowledge among a group of informants (Romney, Weller, and Batchelder 1986). Consensus analysis has two goals: first, to determine the culturally correct answers to questions relative to a particular domain and, second, to evaluate the "cultural competence" of each informant (Borgatti 1993). The first goal is that which is most relevant to our work.

Romney, Weller, and Batchelders' cultural consensus theory is based on the insight that informants who agree with one another about some item of cultural knowledge tend to know more about the domain than informants who disagree with each other. The idea is illustrated in West Africa on manioc classification. Researchers walked 58 women through a manioc garden and asked them to identify the various plants. They found that the more women agreed with each other on the identification of the plant, the more they were likely to know what the plant actually was. In other words, as cultural competence increased, so did cultural consensus (Ryan and

Martinez 1996). As for the Delphi method, a focus group of "specialized" informants is required to conduct these exercises.

Which Method?

The choice among the three approaches presented above should be informed by the concept to be defined. This project requires that the concepts of wealth, poverty, food security, and food vulnerability be defined in their local meaning. Table 4A.1 suggests guidelines to the exploration of those concepts.

Once the meaning of those concepts has been elicited, some additional exploration may be appropriate. For instance, in the normative diet, a rank ordering of essential foods could be obtained through pairwise scoring or contingent valuation. These tools will be reviewed later.

Alternative Methods for Impact Evaluation: The SWOT Analysis

SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis is a common tool used by program managers to elicit and analyze the relative merits and deficiencies of particular activities, and possibilities for improvement. This instrument was initially developed for use by specialists, but it can easily be adapted to an RA setting as its realization is well developed and very straightforward. SWOT analysis is easily explained to participants using a matrix (Figure 4A.2) where the time frame (present/future) is placed on one axis, and evaluations (positive/ negative) are on the other.

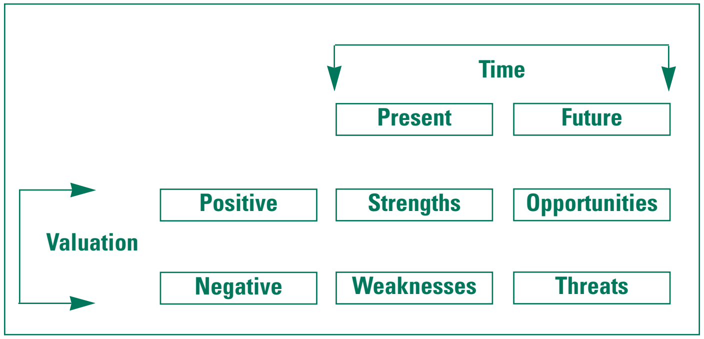
This framework is particularly well-suited to examine the present performance of single development activities for food security (say, credit or technical extension) and evaluate their future implications. Considering the present, what works well and why (strengths) is first examined. Informants work in a brainstorming mode, where all

Table 4A.1 Concepts to define, approaches to use, and outputs to obtain

Concept	Approach	Format/participants	Output
Wealth and poverty	Cultural domain	Focus group/cross section of all villages	List of attributes associated with wealth and poverty in that community
Household configurations	Cultural domain	Focus group/cross section of all villages	List of household forms (extended, nuclear, gender of household heads) and their relative occurrence
Food security	Delphi	Separate focus groups of men, women, project staff	List of attributes associated with food-secure and food-insecure situations; may also include a list of graded responses to food insecurity (to be used as indicators)
Indicator for food security	Cultural domain	Focus group/cross section of all villages	Ordered list of responses to food insecurity
Food vulnerability	Cultural domain	Focus group/cross section of all villagers	List of local livelihood strategies and of threats to these strategies
Normative diets	Delphi or cultural consensus	Focus group/senior women of households	Minimum list of foods and their quantity needed by average adult to lead a healthy life

Source: Compiled by author.

Figure 4A.2 SWOT matrix



Note: SWOT stands for strengths, weaknesses, opportunities, and threats.

comments are welcomed and listed. The same is done with what is not working, and why, in present implementation plans (weaknesses). The examination of future opportunities may refer to ways to improve on present weaknesses, or new initiatives that may be added that would enhance the present strengths. Future threats refer to possible negative impacts of the activity on food security or the emergence of constraints that may impede the continuation of identified strengths or the realization of future opportunities. Programmatic implications naturally follow from these considerations.

Appendix 4B: Impact Evaluation Instruments (Example from Honduras)

Community:____________ **Group:** ____________ **Rapporteur:** ________________

Activities/goals	Impacts			Conditions
	↗Y	↗AA	↗DA	

Appendix 4C: Summary of Impact Evaluation

Sixteen types of interventions were carried out in total (Table 4C.1). Interventions were not always the same from village to village, as (1) the choice of activity was defined by community members themselves; (2) the service provider varied from village to village; and (3) programs were generally directed either at men or at women, and levels of participation varied by gender between communities. Global evaluations of the programs are thus difficult to make, and we can only offer crude measures of the general performance of the activities promoted by PLANDERO in the ten communities. Disaggregating measures by gender, by service provider, by intervention type, and by community can, however, improve the evaluation. The analysis is supported by a review of the reasons invoked by informants as to the reason for success or failure of each activity.

Respondents felt that about one of every three (32.8 percent) of PLANDERO activities improved the food security of their income. This rate of approval differs by gender, with women positively viewing the contribution of activities to food security 41 percent of the time, and men 25 percent of the time. The various dimensions of food security were also rated differently by gender. Overall, 24 percent felt it improved their income, 50 percent felt it improved the local availability of foods, and 25 percent felt it improved their access to food. When contrasted by gender, however, men viewed positively the contribution to income in 16.8 percent of cases; to food availability, in 36.9 percent of cases; and to food access in 23.2 percent of cases, while women viewed positively the contribution to income in 36.6 percent of cases; to food availability in 56.6 percent of cases; and to food access in 29 percent of cases (Table 4C.2).

Table 4C.1 Summary of impact evaluation

Activity	Affected	Villages 1	2	3	4	5	6	7	8	9	10	11	12	Number of communities where activity is deployed, M/F groups	Number of positive impacts, M/F groups	Percent of positive impacts, M/F groups	Mean rate of positive impact across M/F groups
Augment maize production of (wheat)	Y	0/1	0/1	0	0	0	0	0/0	0/1	0/1	0/–	1	1/0	12/11	2/4	17/36	27
	AA	0/1	0/1	1	1	0	1	1/0	1/1	1/1	1/–	1	1/1	12/11	9/9	75/82	79
	DA	0/1	0/1	0	0	0	0	1/0	0/0	0/1	1/–	1	0/1	12/11	3/5	25/45	35
Augment production of beans	Y	–/–	–/–	0	0	0	1	–/–	1/0	–/–	0/–	1	1/1	8/7	4/3	50/43	47
	AA	–/–	–/–	0	1	0	1	–/–	1/1	–/–	0/–	1	1/1	8/7	5/5	63/71	67
	DA	–/–	–/–	0	0	0	1	–/–	1/0	–/–	0/–	1	1/1	8/7	4/3	50/43	47
Augment production of coffee	Y	–/–	–/–	1	–	–	1	–/–	0/–	–/–	–/–	–	–/–	3/2	2/2	67/100	84
	AA	–/–	–/–	1	–	–	1	–/–	0/–	–/–	–/–	–	–/–	3/2	2/2	67/100	84
	DA	–/–	–/–	1	–	–	1	–/–	0/–	–/–	–/–	–	–/–	3/2	2/2	67/100	84
Augment production of horticulturals	Y	–/–	–/–	–	0	0	–	–/–	0/–	–/–	0/–	–	0	5/3	0/1	0/33	18
	AA	–/–	–/–	–	0	0	–	–/–	0/–	–/–	0/–	–	0	5/3	0/1	0/33	18
	DA	–/–	–/–	–	0	0	–	–/–	0/–	–/–	0/–	–	0	5/3	0/1	0/33	18
Diversify production	Y	0/–	–/–	–	–	–	–	–/–	–/–	0/–	–/–	–	–/–	2/–	0/–	0/–	0
	AA	1/–	–/–	–	–	–	–	–/–	–/–	0/–	–/–	–	–/–	2/–	1/–	50/–	50
	DA	0/–	–/–	–	–	–	–	–/–	–/–	0/–	–/–	–	–/–	2/–	0/–	0/–	0
Built conservation infrastructures and agroforestry systems	Y	1/–	–/–	–	–	–	0	0	–/–	0/–	0/–	–	–/–	5/2	1/0	20/0	10
	AA	1/–	–/–	–	–	–	0	0	–/–	1/–	0/–	–	–/–	5/2	2/0	40/0	20
	DA	1/–	–/–	–	–	–	0	0	–/–	0/–	0/–	–	–/–	5/2	1/0	20/0	10
Protect and delimit sources of water	Y	0	–/–	–	–	–	–	–/–	–/–	0	–/–	0	–/–	3/3	0/0	0/0	0
	AA	0	–/–	–	–	–	–	–/–	–/–	0	–/–	0	–/–	3/3	0/0	0/0	0
	DA	0	–/–	–	–	–	–	–/–	–/–	0	–/–	0	–/–	3/3	0/0	0/0	0
Stop slash–and–burn practices	Y	1/–	–/–	–	–	–	–	0	–/–	0/–	–/–	–	–/–	3/1	1/1	33/100	67
	AA	1/–	–/–	–	–	–	–	0	–/–	1/–	–/–	–	–/–	3/1	2/1	67/100	84
	DA	1/–	–/–	–	–	–	–	0	–/–	0/–	–/–	–	–/–	3/1	1/0	33/0	17
Involve primary school in environmental activities	Y	–/1	–/–	–	–	–	–	–/–	–/–	–/0	–/–	–	–/–	–/2	–/1	–/50	50
	AA	–/1	–/–	–	–	–	–	–/–	–/–	–/0	–/–	–	–/–	–/2	–/1	–/50	50
	DA	–/1	–/–	–	–	–	–	–/–	–/–	–/0	–/–	–	–/–	–/2	–/1	–/50	50
Credit education and programs	Y	0	0	1	1	0	1	1/0	–/1	0/1	0	1	1/1	11/12	6/9	55/75	65
	AA	1/1	0	1	1	0	1	1/0	–/1	1/1	1	1	1/1	11/12	9/10	82/83	83
	DA	1/1	0/1	0	1	0	1	1/0	–/0	0/1	0	1	1/1	11/12	6/7	55/58	57
Extension in environmental protection	Y	–/–	0	–	–	–	–	–/–	–/–	–/–	–/–	0	–/–	2/2	0/0	0	0
	AA	–/–	0	–	–	–	–	–/–	–/–	–/–	–/–	0	–/–	2/2	0	0	0
	DA	–/–	0	–	–	–	–	–/–	–/–	–/–	–/–	0	–/–	2/2	0	0	0

Source: Compiled by author from survey data.

Table 4C.1 Summary of impact evaluation (cont.)

Activity	Affected	Villages 1	2	3	4	5	6	7	8	9	10	11	12	Number of communities where activity is deployed, M/F groups	Number of positive impacts, M/F groups	Percent of positive impacts, M/F groups	Mean rate of positive impact across M/F groups
Extension in handling minor species (also value–added production)	Y	–/–	0	0	–	–1	–	0	–/0	–/–	–/0	0	–/–	4/7	1/1	25/14	20
	AA	–/–	0	1	–	–1	–	0	–/0	–/–	–/1	0	–/–	4/7	2/3	50/43	47
	DA	–/–	0	1	–	–1	–	0	–/0	–/–	–/0	0	–/–	4/7	2/2	50/29	40
Improve commercialization	Y	–/–	–/–	–	–	–	0	–/–	0	–/–	–/–	–	–/–	2/2	0/0	0/0	0
	AA	–/–	–/–	–	–	–	0	–/–	0	–/–	–/–	–	–/–	2/2	0/0	0/0	0
	DA	–/–	–/–	–	–	–	0	–/–	0	–/–	–/–	–	–/–	2/2	0/0	0/0	0
Family/school garden	Y	–/–	–/–	–	–	–	–	–/–	–/1	–/–	–/–	–	–/–	–/1	–/1	–/100	100
	AA	–/–	–/–	–	–	–	–	–/–	–/1	–/–	–/–	–	–/–	–/1	–/1	–/100	100
	DA	–/–	–/–	–	–	–	–	–/–	–/0	–/–	–/–	–	–/–	–/1	–/0	–/0	0
Improve women/ youth participation	Y	–/–	–/–	–	–	–	–	–	–	–/0	–/–	–	–/–	–/1	–/0	–/0	0
	AA	–/–	–/–	–	–	–	–	–	–	–/1	–/–	–	–/–	–/1	–/1	–/100	100
	DA	–/–	–/–	–	–	–	–	–	–	–/0	–/–	–	–/–	–/1	–/0	–/0	0
Foment handicraft industries	Y	–/–	–/–	–	–	–	–	–	–	–/0	0/–	–	–/–	1/1	0/0	0/0	0
	AA	–/–	–/–	–	–	–	–	–	–	–/1	0/–	–	–/–	1/1	0/1	0/100	50
	DA	–/–	–/–	–	–	–	–	–	–	–/0	0/–	–	–/–	1/1	0/0	0/0	0

Source: Compiled by author from survey data.

Note: Data are distingued by gender (male/female) when appropriate. Y = income; AA = food access; DA = food availability. In the 12 columns under villages, 0 means "had no positive impact; 1 means "had positive impact"; and – means"no activity was reported." M = male; F = female. "Mean rate of positive impact" is a simple average of the proportions of male and female groups reporting a positive impact.

Key to Villages: 1: El Aguacate; 2: Barrio San Juan; 3: Boca del Monte; 4: La Mohaga; 5: El Moral; 6: Nueva Virtud; 7: Plan El Rancho; 8: El Rosario; 9: San Marcos; 10: Tepezcuintle; 11: El Pinal; 12: Laguna Seca

Table 4C.2 Individuals viewing intervention positively on dimensions of food security, by gender

Dimensions of food security	All informants	Male informants	Female informants
Household income	24.1	16.8	36.6
Availability of food in community	50.4	36.9	56.6
Access of food by household	25.2	23.2	28.5
Improved food security	32	23	40

Source: Compiled by author from survey data.

Endnotes

1. Rapid appraisal (RA) techniques and participatory rural appraisals (PRA) are often thought to be the same: They seek local input using similar techniques and assuming a similar attitude on the part of project staff. There are differences, however. The ultimate goal of PRA is community empowerment. This involves intensive community participation and assumes an open research agenda. This can hardly be done quickly. RA methods, by contrast, are meant to provide researchers with data quickly. RA requires the participation of community members but the research agenda is predefined and the time frame is short. Use of the word "participatory" here is thus in reference to a methodological style rather than an epistemological posture.
2. For the purpose of the exercises described in this manual, a typical team is composed of one "moderator," who explains the activities, channels the interactions, and so on; and one "relator" who takes notes and keeps track of all the information that is provided, including that which does not get transcribed on the final group output. One such team is required for each working group.
3. It is assumed that the situation here is one in which no previous contacts exist and no activities have yet been programmed for that community. The situation will obviously be different if the community graduates from a previous development program, or if development activities have already been defined.
4. There are several definitions of food security that can be found in the literature. USAID for instance includes food utilization (in addition to food availability and food access) as part of the definition of food security, whereas FAO, IFAD and UNDP include only food access and food availability. Since this chapter was done under commission for IFAD, its definition of food security, which includes only access and availability, is used.
5. This method is quite tolerant about choice of respondents: In fact, it is preferable to avoid selecting respondents, as the concept should have as wide a currency as possible among inhabitants of the target village. It is thus best carried on in a workshop setting where all villagers are invited.

5. Constructing Samples for Characterizing Household Food Security and Monitoring and Evaluating Food Security Interventions

Calogero Carletto

Introduction

Reliable information on household food security is a prerequisite for the accurate and effective design, monitoring, and evaluation of development projects. In marginalized areas, where many development agencies work, this information is often either not available or grossly out-of-date. But collecting data is not a costless exercise. This chapter discusses how random sampling techniques—methods that use some mechanism involving chance to determine which farms, households, or individuals are to be studied—can economize on the costs of gathering information while increasing the likelihood that it will be both accurate and available in a timely fashion.

The chapter begins with a brief explanation of why random sampling techniques are a powerful means of obtaining information on household characteristics such as food security. It then takes the reader through a step-by-step process of constructing a random sample. Having outlined these issues, a worked example is then presented. Readers interested in pursuing the issues raised in this guide are encouraged to consult Bernard (1988), Casley and Lury (1987), Casley and Kumar (1988), Devereux and Hoddinott (1992), and Newbold (1988). More technical discussions are found in Kish (1965) and Cochran (1977).

Why Random Samples?

Random Samples Rather than Censuses

One alternative to a random sample would be to obtain information on all observations in a population census or a census of agriculture. The advantage of a census is that it seemingly provides an accurate "snapshot" of the population at a particular moment in time. It also ensures that numerically small groups, which might be missed in a survey, are counted. Censuses are characterized by (1) individual enumeration (each unit of observation, say farm household, is measured separately); (2) universality within a defined territory or domain (information is obtained on everyone in a certain area); and (3) simultaneity (everyone is interviewed at the same point in time). The key criterion is simultaneity. The census should be conducted within a short and well-defined period of time to reduce omissions and duplications.

There are a number of drawbacks to conducting a census. First, it is usually much more expensive than conducting a survey. (This is not true, of course, where the population is very small.) Second, the processing and cleaning of a census is enormously time consuming. Further, a smaller sample allows the researcher to devote extra effort to ensure the information obtained is accurate. The gains from a smaller, more accurate survey could well outweigh the benefits of obtaining less accurate information on a much larger group. Finally, many topics, such as those involving detailed transactions of individuals or firms, require extensive interviews or observations that cannot be carried out in a census. So issues of cost, time, precision, and quantity of data all suggest that a survey is preferred over a census.

There is a further reason. Censuses are unnecessary. You can learn all you need to know about a given population with a random sample of that population. This is referred to as inference. You draw a sample of a certain number of observations from a given population then calculate parameters of interest such as means and proportions that, by inference, represent the characteristics of the underlying population.

Random Versus Nonrandom Sampling

It is not necessary to obtain information on all units of observation. Is it necessary, however, to choose those households or farms to be studied in a random, or probabilistic, fashion? Why not use nonrandom, or nonprobabilistic, methods instead?

Nonprobabilistic methods are those in which the analyst consciously chooses who will be interviewed. Examples of these include the following. One is accidental sampling. This involves interviewing respondents as they are found, for example, walking down a track or road and interviewing whoever you happen to meet. A second is quota sampling. Here, enumerators are instructed to contact a specified number of observations possessing certain characteristics (for example, 15 farms with no livestock; 10 farms with 1 to 3 head; 5 farms with more than 3 head of cattle). These quotas are assigned on the basis of what is known about the underlying population. However, the actual selection of observations is left up to the enumerator. A third method is purposive sampling. Here, individual units of study are chosen on the basis of some judgment criteria. Suppose you want to learn about long-term processes of environmental change in a rural area. To obtain this information, you could choose a sample of "wise old men." A fourth method is referred to as networking. Here, you find one person to interview and ask them to name others who are also suitable candidates, given the topic of interest. If the population is small, this can be a useful means of building up a sample. However, in larger populations every person or unit of observation does not have an equal chance of being sampled—that is, the sample selected is not representative of the underlying population. However, networked samples are very useful when exploring networks (you want to find people who know each other) or when dealing with hard-to-find groups.

Under a number of circumstances, nonprobabilistic sampling methods are appropriate. For example, if the population is homogeneous ("describe one unit of observation and you describe them all"), these methods produce information identical to that derived from probabilistic techniques. They are also appropriate if there is no intention to extrapolate the results to the larger population (for example, where the objective is to describe a village in general terms rather than obtain a statistically representative picture). Finally, such methods are useful where a sampling frame is unavailable or too costly to obtain.

But there are also significant drawbacks to these methods. Statements made on the basis of observations found in these ways must be limited to the sample itself—it is not possible to make legitimate inferences about the wider population. Further, standard statistical techniques—such as comparing means of two groups—cannot be used either. For these reasons, the use of nonprobabilistic methods by development practitioners is strongly discouraged. They should only be used where probabilistic methods are infeasible.

Steps in Constructing a Random Sample

There are five steps involved in constructing a random sample: (1) determining the sample unit, (2) determining the "universe," (3) constructing a sampling frame, (4) deciding on the sample size, and (5) choosing the sample. Although these are discussed sequentially, it should be noted that it is often necessary to iterate back and forth among these. So, for example, practical considerations associated with choosing a sample may have an impact on the manner in which the sample frame is constructed and the calculation of the sample size. A glossary of sampling terms is provided in Box 5.1.

Box 5.1 A glossary of sampling terms

Universe	The location or population or group that the analysis seeks to describe.
Sampling units	The unit of observation of the study, such as farms, households, individuals, and so on.
Sampling frame	The list of sampling units. It must contain all units within the universe.
Self-weighting samples	A sample in which all units have an equal probability of selection.
Sampling fractions	The ratio between the sample size and the population size. Also called the selection probability.
Domain	A part of the population for which separate estimates are sought. Examples are farms of a certain size, individuals of a particular age group.
Cluster	The aggregation of sampling units, often based on geographic proximity. Examples are a village or a section of village.
Take	The number of sampling units drawn from a selected cluster.

Source: Compiled by author.

Determining the Sampling Unit

The appropriate sampling unit—or unit of observation—is guided largely by the objectives of the survey and the project. For example, where a project seeks to increase farm yields, the relevant sampling unit for evaluation purposes would be the farm household. If the objective was to improve the nutritional status of children under five, the relevant unit of observation would be children in that age bracket. What is important here is that the definition of the sampling unit should be unambiguous and conform to local understanding and acceptance. The most common ultimate sampling unit in multipurpose socioeconomic studies is the household, even if individual-specific estimates are sought. In some countries, there may be a generally accepted definition of what constitutes a household—for example, the definition adopted by the central statistical office. Even where such a definition exists, it should be validated locally before proceeding with the listing exercise.

Determining the Universe

The "universe" is the location or population or group that the study seeks to describe. Again, this is likely to be determined by the objectives of the project. If the project is located in, say, western Honduras, then western Honduras would be the location of the study. However, it is not always practical to survey the entire location. The discussion below on "choosing the sample" and the worked example from Malawi illustrate solutions that are available when this problem arises.

Constructing a Sample Frame

The use of probabilistic methods to select a sample requires a sample frame.

The frame for a sample is a list of the units in the population (or universe) from which the units that will be enumerated in the sample area are selected. It may be an actual list, a set of index cards, a map, or data stored in a computer. The frame is a set of physical materials (census statistics, maps, lists, directories, records) that enables us to take hold of the universe piece by piece (Casley and Lury 1987, 52.)

Examples of lists that can be used as sample frames include lists of administrative areas, census materials, ordinance survey maps, tax listings, land registries, and lists of project beneficiaries. In practice, there are a number of dangers when working with such materials. Take, for example, a list of households. First, the frame may be inaccurate. This could result from errors in recording information—names might have been misspelled, adults were listed as children, households contain people who are not recorded, and so on. Alternatively, these errors might have occurred because the information was collected from neighbors because household members were absent or unavailable when the frame was created. Second, the frame might be incomplete. Households or groups of households may have been omitted. This might have occurred because the frame is out-of-date (for example, households have subdivided or migrated in or out) or because of poor enumeration when the frame was created. There might have been difficulties in determining the location of boundaries, with the result that certain households were missed. Third, there might be duplication. Some households are included twice, possibly because (1) the lists were compiled by more than one person; (2) there is confusion over names; or (3) disputes over land claims exist. Devereux and Hoddinott (1992) provide a good example of some of these problems in this description of surveying households in northern Ghana.

"When I first arrived in my chosen village of Pusiga I introduced myself to the subchiefs in the two sections, Terago and Tesnatinga (or Teshie), in which I planned to work. These sub-chiefs had recently compiled lists of households for their sections, which were used by the District Administration to distribute small quantities of government food aid (following the two successive poor harvests mentioned above). Had these lists been compiled for an unpopular purpose such as tax collection, I would have had reservations about their accuracy. But since everybody had an incentive to register for food aid, I decided to use the sub-chiefs' lists as a basis for household enumeration.

Nonetheless, these lists were inaccurate in several respects. . . . Over-reporting occurred mainly in large, complex compounds, and typically took the form of young men claiming to be separate households when they were, in fact, still farming with their brothers or father. The explanation for this was simple. When the household lists had been drawn up, local residents were well aware that the purpose was to distribute food aid. People in large compounds reasoned that if each household was to receive free food, it was to their advantage to exaggerate the number of separate households in their compound. When I made my first round of interviews, the expectation that I would be bringing some kind of free or subsidized assistance to the village was high, and overreporting was standard practice. During the year I gradually discovered which compounds had overreported household numbers, and simply crossed them off my list. (A clear indicator was when I asked several 'household heads' in a compound about planting, harvests, and asset ownership, and received identical or near-identical figures—since they were each listing, in fact, the same (joint) production and assets.) . . .

Under-reporting of households occurred most commonly with old women, especially widows. Although most old widows are looked after by a son or son-in-law, this is not always the case, and some old women constitute separate households, either because they insist on retaining their independence by farming their own land, or because they have been cast adrift to fend for themselves. In my sampling frame there were three such single-person 'households,' one in the first category and two in the second, all of which I missed until it was too late to incorporate them in the lists of households from which my samples were randomly selected.

The reason why these widows were missed is to be found in the local conceptualisation of a household, which corresponds broadly to the Western notion of a 'production-consumption unit.' A man is said to constitute a separate household if he is 'farming separately' (from his father and brothers) and 'feeding himself' (and his wives and children)—that is, both a production entity and a consumption entity. The two widows living on their own were virtually beggars, being too infirm to work and having no one to help them with farming. In fact, they were dependent on handouts from relatives and neighbours. So they did not strictly qualify as households in terms of the local definition because they were neither 'farming separately' nor 'feeding themselves'" Deveraux and Hoddinot (1992, 50–53).

It follows that survey designers should always plan to have any existing list checked. In monitoring and evaluation exercises, the population under study, or at least a domain of it, is generally composed of the beneficiaries of a certain project or program. In many instances, the lists of project beneficiaries are readily available with the project management. However, even in these apparently favorable conditions, it is imperative to check these lists for inconsistencies, omissions, and duplications. By no means should their accuracy be taken for granted.

Where no such listing of households, or units of observation, is available, or where such lists are so outdated or inaccurate as to be useless, two possibilities remain. These are (1) to create a list or (2) to derive a sample without a frame. These are discussed in turn.

Creating a sample frame can be a time-consuming and expensive exercise. For this reason, there may be practical advantages associated with using multistage sampling (described below) or restricting the universe to be studied. For example, an evaluation of a project might be limited to certain localities rather than all areas in which a project has operated. It is important to note that by adopting such strategies, probabilistic samples are representative of a restricted universe and as such any extrapolation of the results should be confined to it.

One approach is to start with a list, even one that is known to be inaccurate. For example, in northern Mali, survey work began with lists of households that had been compiled several years previously for the distribution of food aid. The survey team, accompanied by village leaders, walked through the villages matching names on the list to households, adding new names, and deleting those no longer resident in the village.

Where even rudimentary lists are unavailable, maps can be used as a starting point. A first step in area sampling may involve the use of a map providing a graphical representation of the universe, for example, a region or a province. Using easily identifiable natural boundaries, the map can then be partitioned into approximately equal-sized segments. Once all the segments/villages have been delimited and some chosen, sketch maps can be easily produced in a relatively short time without need of much expertise. Of course, the amount of time and resources going into this mapping exercise should be suited to the objective at hand. In most cases, very rough sketches describing the main roads and pathways and some landmarks (for example, a church, a mosque, a borehole, a river and so on) clearly delimiting different sub-areas of the segment/village could suffice. In most cases, however, the inclusion in the sketch of the individual domiciles, properly numbered, may be necessary. As in the case of working with listings, it is important to verify that no area or sampling unit of the universe has been missed and that no overlapping occurs between different maps, since this would obviously result in unequal probability of selection for the elements of the chosen population.

There may be instances where it is simply not feasible to construct a sampling frame. In such circumstances, the following two-stage technique—EPI Cluster Survey Design, developed originally to monitor and evaluate the Expanded Programme on Immunization (EPI)—can be used.

The original design, used for the monitoring of immunization coverage of children within a target age (generally 12–23 months), involves the selection of 30 primary sampling units or clusters (villages or other types of area units), and the subsequent drawing of seven ultimate units (children) from each cluster, for a total sample size of 210 children. The clusters are selected from a comprehensive list of villages or area units with probability proportional to estimated cluster size. Census information and administrative records may be used to generate the list containing the estimated size of the cluster. The second-stage selection of seven children in each cluster was originally envisioned as a random selection from a list of children in the target age living within the cluster. However, enumeration difficulties have led to the development of more simplified procedures. A commonly used variant of the original scheme suggests choosing a random direction from a central point in the village/area unit by spinning a pen or bottle. Only the households along this direction up to the edge of the cluster are enumerated, and one is chosen at random. Starting from the chosen households, and along the direction line, seven adjacent households with children in the target age are selected and interviewed.

A plethora of variants have been tried in recent years to partly overcome some of the limitations associated with the standard design. Choosing seven adjacent households in the case of a restricted target group (such as children between 12–23 months) may actually result in a quite spread-out sample within the cluster. In another circumstance in which eligibility criteria are not so stringent, the selection is likely to identify a highly concentrated conglomerate of households. Under the plausible assumption that adjacent households exhibit very similar socioeconomic characteristics, it is evident why the standard design does not perform well in multi-indicator socioeconomic surveys.

With the goal of selecting more heterogeneous elements within the cluster, a possible variant to the standard design would be to select the third, fourth, or fifth household, starting from a central (or randomly chosen) location after a direction is picked. From this last selected household, you repeat the procedure and proceed in a random-walk fashion until the quota is met. Alternatively, the village could be split into smaller areas and from the center of each unit (or any randomly chosen point), a direction picked, and the *n*th household meeting the eligibility criteria interviewed.

Deciding on the Sample Size

Calculating sample sizes is one of the most technically demanding aspects of survey design. Although a number of software packages—such as Epi-Info and STATA—automate these calculations, it is still necessary to understand what information is required in order to run these routines. This subsection provides an overview of these issues.

Abstracting from practical issues such as the time and resources available to undertake a survey, a decision regarding sample size is strongly linked to the required level of precision in the variables we seek to measure. Precision—or sampling error—is described in terms of a margin of error and a confidence level. For example, you might want to estimate sample maize yields within 3 percent of the true mean (the mean we would obtain if we measured all maize yields). This statement implies that if you were to take 100 samples, you would expect that the sample means would be within 3 percent of the true mean at least 95 times. Three other factors will also play a role. One is the distribution of the variable of interest. If maize yields are identical across all households, then you would only need to sample one household in order to determine the average level of maize yields. By contrast, more dispersed distributions require a larger sample size. Second, sample sizes are affected by the particular sampling design chosen. Multistage designs require larger samples than single-stage designs in order to achieve the same degree of precision. Third, increasing the number of variables that you seek to estimate may also affect the sample size needed to attain a certain degree of precision.

Finally, there is a widespread but erroneous belief that sample size depends on the size of the population and therefore on the sampling fraction. The size of the population only marginally affects the precision of the estimate. The precision of the estimate is directly related to the absolute size of the sample, but much less so to the sampling fraction. A sample of 100 units drawn from a population of 1,000 (sampling fraction 10 percent) is highly unlikely to produce more precise estimates than a sample of 200 from a population of 10,000 (sampling fraction 2 percent).

Choosing the Sample

Armed with a suitable sample frame that lists units of study and knowledge regarding the desired sample size, the last step is to select the sample in a random or probabilistic fashion. There are four types of probabilistic methods: systematic, simple random sampling, stratified, and multistage.

A relatively straightforward method of selection is systematic sampling, where draws are made at fixed intervals through the list starting from a random unit. For example, suppose you want to extract the same sample of 10 households from the list of 150 households. You randomly select a number between 1 and 15 (150 divided by 10) and, starting from that unit, select every 15th household. If 5 were the randomly selected number, then the sample would be composed of households 5, 20, 35, 50, 65, 80, 95, 110, 125, and 140.

Note that in addition to being a random selection method, this method has another advantage when the list is ordered on the basis of some feature related to the variable of interest. Suppose you want to estimate crop yield, and the list is ordered based on farm-size class; then systematic selection would guarantee that a wider spectrum of farm-size classes are represented in the sample. Following this systematic method, you can be almost certain that the first sample element (household 5) belongs to a different class of, say, element 8 (household 110) or 10 (household 140). Set against this advantage is a potential danger. If there is some subtle, difficult-to-observe ordering of the sample (resulting, for example, in small farms never having numbers ending in zero or five), the observations drawn will not be a random sampling of the population.

A second method is systematic random sampling (SRS). A simple illustration of this is the following. Write all the farm identification numbers on individual slips of paper and throw these in a hat. Shake the hat vigorously. Pick out the number of farms you want to interview and that is your sample. However, in large populations this is a rather tedious operation (and might require a very large hat!). An alternative method is to use a table of random numbers.

There is a potential weakness with both systematic and systematic random sampling. Suppose you are drawing a sample of 100 farms from a population of 1,000. You know from the census that 30 percent of these have more than 10 acres of land, so the sample should contain 30 such farms. However, this is only true on average. Though the likelihood is high that the sample will contain 30 large farms, it is also possible that it will contain 20, 25, or 40. Suppose it contains only 15 such farms. Other things being equal, if larger farms have better access to formal-sector credit than smaller farms, and given that the larger farms have been underrepresented, you might feel that inferences regarding credit will not be reliable. One tempting possibility would be to pick two or three other samples and choose the one you thought was most representative. The difficulty with this approach is that the sampling procedure being used—the population is sampled until you find a sample you like—can no longer be justified and the results are no longer suitable for the purposes of inference.

There is a solution to problems such as these: random stratified sampling. The first step is to divide the population into groups or strata. Here, the division would be between the 300 large farms and the 700 smaller ones. Using the random number method, select 10 percent of farms in each category, so the resultant sample contains 30 large farms and 70 small ones. The proportions in the sample are identical to those in the underlying population.

Random stratified sampling is an attractive means of obtaining a sample. However, it is helpful to note two potential problems. First, the relevant stratification variables must be known in advance. Second, you must know the underlying population proportions of each stratum. Addressing these problems requires additional information on each unit of observation. For example, lists of farms may contain only the name of the household head. A short survey may be necessary to obtain information to stratify and this may be too time-consuming or expensive.

A fourth form of sampling is multistage or cluster sampling. Whenever the universe from which you want to draw the sample is geographically spread out, single-stage procedures such as SRS or systematic sampling may not be logistically feasible, since they are both likely to generate equally dispersed samples. The necessity to lower transportation and organizational costs, as well as reduce nonsampling errors (enumerators working on a large area may be more difficult to supervise, increasing the likelihood for errors), suggests that a multistage design may be more appropriate. In addition, multistage designs can produce substantial savings in terms of time and financial resources that must be allocated to the listing operations.

A two-stage design would generally call for the selection of geographically delimited nonoverlapping primary sampling units (PSU), also known as clusters (examples of clusters are a region, a district, a village), the selection of a limited number of clusters, and within each cluster, the random selection of a certain number of ultimate sampling units. Given that a two-stage design is chosen, a number of issues arise. How do you select the clusters from the universe? How many clusters do you select? How many ultimate units do you draw from each cluster?

The way clusters are selected depends primarily on the availability and accuracy of a complete sample frame. In the simplest case scenario in which such a list is available and the clusters are of equal size, we can select a number of them using simple random sampling and, within each, draw an equal number of ultimate units.

A Worked Example

This example outlines how a random sample of farmers was obtained in order to assess the impact of two projects directed toward smallholders in Malawi. As is discussed in Chapter 7, it was necessary to survey participants in both projects, as well as households enrolled in neither ("control households"). The example illustrates practical difficulties encountered in sampling, the solutions adopted, as well as the time requirements of the different steps.

Selecting the Sampling Unit

Both projects targeted smallholder farmers. Consequently, the sampling unit was a smallholder farm household, classified using a local definition as a rural household with less than 10 hectares of land.

Selecting the Universe

The next step was to select the area(s) for the data collection. Based on a classification by the Ministry of Agriculture, the country of Malawi is divided into three regions (North, Central, and South), further divided into extension planning areas (EPAs). Although it would have been ideal to work in all three regions, time and budgetary constraints made it necessary to restrict the survey to a single region. Field visits conducted in the regions (one or two days for each visit, combined with extensive talks with key informants) revealed significant differences between these regions. For this reason, a random selection of one region was not appropriate. Instead, it was decided to select an EPA in Central region. To facilitate the contrast of the two projects and rule out differences in location-specific features (or having to control for them during data analysis), it was decided to select an EPA in which both projects were active. This restricted the choices to a pool of only two EPAs with very similar characteristics; one EPA was randomly selected. An implication of this decision was that it was not possible to extrapolate any findings from this region to the whole country.

Constructing the Sampling Frame and Selecting the Sample

Given that the objective of the study was to compare the two projects against each other, and against the control group, it was necessary to sample households in both projects as well as households in neither project. These three groups constitute separate "domains." (Technically, the universe—the EPA—was stratified by domain.) One way of doing this would have been to enumerate all households in this area and select households from each domain in proportion to their number in the EPA. However, given the relatively low coverage of both projects, this technique would have led to the selection of an insufficient number of observations among the two beneficiary groups. Since the main objective of the study was to compare these groups and not to extrapolate the group or domain estimates to the EPA as a whole, the research team chose to select an approximately equal number of observations in each of the three strata, namely the food security and agriculture development project beneficiary group and the control group.

Smallholder farmers belonging to either project were organized into clubs of variable size between 10 and 30 households. The club was selected as the primary sample unit. The lists of clubs belonging to each project were available with the project management units. But because membership in these groups changed radically over relatively short periods of time, these lists were not considered reliable. Further investigation (one or two days talking to key informants) revealed the existence of several such lists. In some instances, a list would differ from the others quite substantially. We spent several days trying to reconcile the different sources in an attempt to come up with a unique list that reflected actual project membership.

The first step in the verification process was to clearly define membership for each project. Given the objective of the stratification (to enhance the group contrast and measure project impact within each domain), a club was considered a beneficiary of project A if it had been active within the project for at least two seasons and it had never belonged to project B. "Active" meant that it had produced and sold tobacco in both seasons and had participated in the project's activities. Based on the definition, some clubs were excluded from the list, either because of dual membership or because they had not produced and sold any tobacco.

In the case of the food security project, determining membership was slightly easier since it could be related to access to the credit package being disbursed by the project. A club was considered a food security beneficiary if it had received a full or partial credit package in both of the last two seasons. The main difficulty with this group was represented by the common practice for members to use different names in joining the club. Local key informants—field assistants and village headman—helped screen "hidden" duplications.

Once the research team accounted for these sources of omissions and duplications, they ended up with a list of 14 clubs for the food security project and 71 for the agricultural development project. The next step was to enumerate all the members of each club. These lists were not available. The only information readily accessible was the total number of members at the year of club formation. Given the dynamic nature of membership, these figures were not considered reliable. Therefore, enumeration of the clubs was deemed necessary. To reduce the amount of time necessary for this operation, it was decided to select only a limited number of clubs from the agricultural development list. To this end, 30 agricultural development clubs were selected, using a fixed probability of selection.

Once enumeration for these 30 clubs had been completed, because of the variable size of the clusters the research team drew from each cluster a number of households proportional to the size of the cluster. The procedure resulted in a self-weighting sample within the agriculture development domain. Due to the already limited number of food security clubs eligible for inclusion in the sample, a full census was considered appropriate for this domain.

The selection of the control group called for a different methodology altogether. Available census data were more than 10 years old, and hence suspect. Alternative administrative records were not available. Tight time constraints made complete enumeration of the selected villages infeasible. In addition, to reduce transportation costs and to avoid selecting households from villages where neither project was active, it was decided to select a control household for every other beneficiary household in the village in which the beneficiary household resided. One complication was that the exact village of residence of the beneficiary household was not known until the household was actually visited, so enumeration of selected villages in advance was not possible. In addition, time constraints would not have allowed for it. For the selection of the control households, a variant of the EPI cluster design was used.

Once the research team visited a village in which selected beneficiary households lived, they randomly selected one nonbeneficiary household for every other beneficiary household in the sample. For example, assume that a total of eight beneficiary households belong to village *x*. A total of four nonbeneficiary households were to be drawn from this village. The first step was to roughly sketch the village to locate a central point. From this central point, a team of enumerators, jointly with a supervisor, chose a random direction by spinning a pen on a flat surface. Once a direction was selected, the enumerators were asked to follow that direction and, starting from the 4th dwelling, interview the first household that met the eligibility criteria to belong to the group, that is, they owned less than 10 hectares of land and they had never belonged to either the food security or the agriculture development project. If the same team had been assigned another control household, the supervisor would again spin the pen in front of this first selected household, choose a new direction and, starting from the 4th dwelling, identify the next household to be interviewed along this "random walk." If, instead, another team was required to select an additional control household in the same village, the random walk would start again from the center of the village by randomly choosing a new direction.

One of the potential problems with this variant of this selection design is that it tends to underrepresent households located in more remote areas from the village center. To partly prevent this problem, bigger villages were often divided into sub-areas and a center chosen in the sub-area in which the selected beneficiary household fell. Another potential problem was that the selection must follow a natural path, restricting the number of options in terms of the direction an enumerator can take from the central point. Whenever possible, the enumerators were instructed to cut through fields and follow as closely as possible the direction chosen.

Calculating the Sample Size

Sample size calculations took into account the degree of precision required, statistical power, design effects, and estimated nonresponse rates. A total of 202 households per stratum, for a total sample size of approximately 600 households, were pursued. The value corresponds to a prevalence of about 0.7, a difference in the magnitude of 0.15, a one-tailed statistical significance of 95 percent, a statistical power of 80 percent, and a design effect of 2.

6. Targeting: Principles and Practice

John Hoddinott

Introduction

Targeting refers to the practice of limiting access to an intervention to a select group of individuals. Generally, this can be accomplished by: explicitly applying criteria for participation that include some groups, but exclude others (variants of this are described as categorical targeting and individual assessment); allowing, in principle, anyone to participate but setting up the intervention in such a fashion so as to discourage participation by certain groups (typically described as self-selection); or by some combination of the two. It is widely praised as an attempt to reach the poorest of the poor, yet it is not always straightforward to implement. A poorly targeted intervention can be more costly and less effective than one that is randomly allocated or made available to all households. To avoid costly mistakes, development practitioners must understand the principles and practice of targeting.

This chapter considers three principles underlying targeting:

- Targeting should never be undertaken for its own sake, but should be assessed against a benchmark, such as its impact on reducing the severity of food insecurity.
- Targeting is not costless. It is effective only when the benefits associated with additional reductions in food insecurity outweigh the additional costs associated with doing so.
- Where resources are limited, there is a strong case for categorical targeting, for example by using geographical criteria. However, even in this case, regional rankings can be very sensitive to the criteria used in the identification process. Where resources are limited, the case for individual assessment is considerably weaker.

The Principles of Targeting

Defining the Objective

Many development agencies seek to improve household food security, which is generally defined as adequate access to food at all times, throughout the year, and from year to year. Suppose this general definition is specified more narrowly. Specifically, a hypothetical female person is food-secure if the number of calories available for her to eat exceeds her requirements. If caloric availability is less than nutritional requirements, she is described as food-insecure. Accordingly, it is tempting to assume that the objective of targeting is to produce the greatest decrease in the percentage of individuals who are food-insecure.

Unfortunately, matters are not quite so simple. Consider Figure 6.1. The horizontal axis is a ranking of individuals from least to most food-secure. The vertical axis shows individual caloric availability. The horizontal line indicates requirements. Note that the number of calories available to person A is just below her requirements, whereas caloric availability for person B is significantly below her requirements. Suppose enough calories were "transferred" from B to A so that A can now meet her requirements. The measure of food insecurity—percentage insecure—would register an improvement, even though the poorest person has been made worse off. This is presumably not the intention of interventions designed to reduce food insecurity.

Table 6.1 Example of data necessary for calculating P0, P1, and P2

Person	Daily caloric availability	Food-insecure (Yes/No)	Food insecurity gap (if food-insecure, requirement–availability)	Severity of food insecurity (if food-insecure: equals gap squared)
1	2,325	No (=0)	0	0
2	1,900	Yes (=1)	2,200–1,900 = 300	90,000
3	2,100	Yes (=1)	100	10,000
4	1,700	Yes (=1)	500	250,000
5	2,100	No (=0)	0	0
Sum	—	3	900	350,000

Source: Compiled by author.

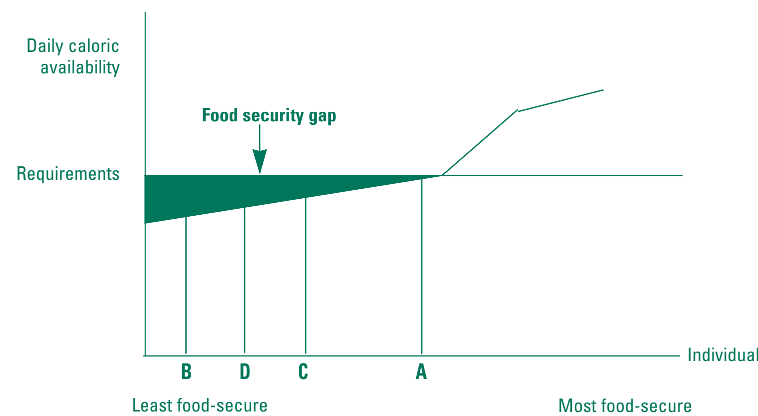
An alternative way of measuring food insecurity might be in terms of a food-insecurity gap. This can be thought of as the total amount of increase in food security needed to eliminate food insecurity among all food-insecure households. In the example above, this would be calculated by adding up the caloric shortfalls of all individuals for whom availability was less than requirements—the shaded area in Figure 6.1. This measure shows the folly of using the percentage measure. In the example above, although the percentage of food-insecure individuals falls, the food-insecurity gap increases. However, consider a second example. The number of calories available to person C is below her requirements. Caloric availability for person D is even lower than C's. "Transferring" a small amount of calories from D to C causes both individuals to remain food-insecure. The percentage measure would remain unchanged as would the food-insecurity gap. However, the most food-insecure person is now even more food-insecure and this is not being captured in either measure. One way of resolving this would be to apply more weight to a reduction in food insecurity among the most food-insecure individuals. Such a measure explicitly emphasizes the severity of food insecurity.

Now consider the following formula:

$$P(\alpha) = (1/n)\sum_{i=1}^{q} [(z - y_i)/z]^{\alpha}, \quad \textbf{(6.1)}$$

where n is the number of individuals; y_i is the measure of food security for the ith person; z represents the cutoff between food security and insecurity (expressed here in terms of caloric requirements); q is the number of food-insecure individuals; and α is the weight attached to the severity of food insecurity.

Figure 6.1 Stylized distribution of food security



Source: Devised by author.

Giving no weight to the severity of food insecurity is equivalent to assuming that $\alpha = 0$. The formula collapses to $P(0) = q / n$, or the percentage measure. This is also called the head-count ratio.

Giving equal weight to the severity of food insecurity among all food-insecure households is equivalent to assuming that $\alpha = 1$. Summing the numerator gives the food-insecurity gap; dividing this by z expresses this figure as a ratio.

Giving more weight to the severity of food insecurity among the most food-insecure households is equivalent to assuming that $\alpha > 1$. A common approach in the poverty literature is to set $\alpha = 2$, yielding

$$P(2) = (1/n)\sum_{i=1}^{q} [(z - y_i)/z]^2. \quad \textbf{(6.2)}$$

Although this formula is fairly straightforward, it can look a little intimidating for someone who has not used it before. For this reason, it is helpful to work through the following example (Table 6.1). Consider caloric availability for five people. Caloric requirements are assumed to be 2,200 calories per day.

Recall that the formula is

$$P(\alpha) = (1/n)\sum_{i=1}^{q} [(z - y_i)/z]^a, \quad \textbf{(6.3)}$$

where n is the number of individuals; y_i is the measure of food security for the ith person; z represents the cutoff between food security and insecurity; q is the number of food-insecure individuals; and α is the weight attached to the severity of food insecurity.

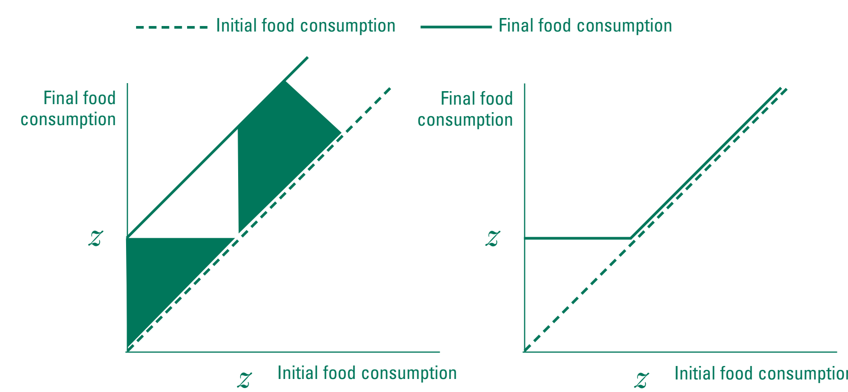
Here, there are three food-insecure people, so P0 (percentage of food-insecure people) = 3/5 = 0.6. The food-insecurity gap, P1, is (1/5) (900/2,200) = 0.08. Finally, the severity of food insecurity, P2, is (1/5) (900/2,200)2 = 0.03.

Which measure should be used when considering the impact of targeting an intervention so as to reduce poverty or food insecurity? If the objective is merely to reduce the percentage of poor, or food-insecure people, than P(0) is the correct measure. If the objective is to reach out to the poorest of the poor, then P(2) is the correct metric.

The Benefits and Costs of Targeting

The basic case for targeting is tantalizingly simple. As above, food security is defined in terms of their being enough calories available for individuals to satisfy there requirements. Using survey data, food acquisition is graphed, ordering the sample from worst to best-off. This initial ordering is represented by a dashed line in both panels of Figure 6.2. In the left-hand panel, a uniform transfer of calories of amount z is given to every person. By doing so, every person meets his or her minimum caloric requirements. In the right-hand panel, anyone with caloric consumption less than z is given a transfer sufficiently large so as to bring initial consumption plus transfer up to minimum requirements. This achieves the same objective but at far less cost. The uniform transfer is plagued by two sources of excessive expenditure: leakages to the nonpoor (represented by the black quadrilateral) and payments to the poor in excess of their needs (represented by the empty triangle).

Figure 6.2 The benefits of targeting



Source: Devised by author.

The case for targeting is complicated by several factors. First, targeting is not costless—it imposes administrative costs that reduce the amount of money available for the actual intervention. These costs will vary with the degree, or fineness, of targeting. One might imagine that there are certain fixed costs associated with targeting. Initial targeting, say, on the basis of geography, may be relatively costless. As targeting moves below a certain geographical level (say, the district) from villages to households and individuals, it becomes increasingly costly. Second, when interventions are targeted, there is a real possibility that some food-insecure households will be missed and some food-secure households will benefit. This can be described as errors of inclusion and exclusion. An error of inclusion is one in which an intervention reaches individuals who were not intended to be beneficiaries. An error of exclusion occurs when intended beneficiaries are not able or permitted to participate in the intervention. Table 6.2 provides an illustration.

Table 6.2 Errors of inclusion and exclusion

Participation	Food-insecure	Food-secure	Total
Participate in intervention	Success 45	Inclusion error 20	65
Do not participate	Exclusion error 15	Success 20	35
Total	60	40	100

Source: Compiled by author.

There are four groups in Table 6.2. An indication of successful targeting is when food-insecure households participate in the intervention and food-secure households do not participate. The food-secure who participate, 20 percent of the population, are counted as an error of inclusion. The food-insecure who do not participate, 15 percent of the population, are counted as an error of exclusion.

An alternative way of looking at this phenomenon involves calculating leakage and undercoverage rates. Leakage is calculated by looking at program participants—those found in the top row of Table 6.2. The number of food-secure beneficiaries is divided by the total number of participants—20/65, yielding a leakage rate of about 30 percent. Undercoverage is calculated by looking at those who should be participants in the intervention but are not—those found in the second row of the second column of Table 6.2—relative to the total number of potential beneficiaries. The number is divided by the total number of food-insecure households—15/60, yielding an undercoverage rate of 25 percent.

All other things being equal, lower leakage (inclusion error) is preferable to higher leakage. Lower undercoverage (exclusion error) is preferable to higher exclusion error. Why do these errors exist? Some undercoverage may be due to factors such as lack of knowledge that the intervention exists or the presence of constraints (say, catastrophic illness or sudden death, which reduces household labor supply) that make it impossible for an eligible household to participate. Some eligible households may decide that the benefits associated with participation do not outweigh the costs associated with doing so. Some leakage may occur due to faulty project design or implementation.

Two additional factors that affect leakage and undercoverage rates are the indicators used to screen participants and the resources available to fund participation. In order to focus solely on these, suppose that none of the reasons for inclusion or exclusion listed above are applicable. There are 100 households in the sample, of which 33 are food-insecure (Table 6.3). Consider, as a baseline, a scenario in which there are enough resources to provide this intervention to exactly 33 households. In the absence of any further information on these households, participation is by random draw.

Table 6.3 Errors of inclusion and exclusion under random draw

Participation	Food-insecure	Food-secure	Total
Participate in intervention	Success 11	Inclusion error 22	33
Do not participate	Exclusion error 22	Success 45	67
Total	33	67	100

Source: Compiled by author.

Leakage is 67 percent (22/33), as is undercoverage. In fact, leakage is constant no matter how many households participate. Undercoverage falls monotonically from 100 percent—when all households are excluded—to zero when all households are included.

Now consider the case where you can target on the basis of an indicator that perfectly captures household food security. This generates Table 6.4. Here, both errors of inclusion and exclusion are zero, as are the measures of leakage and undercoverage. Now suppose all households are permitted to participate. Undercoverage remains at zero, but leakage would rise to 67 percent. Conversely, if no household were permitted to participate, leakage would be zero, but undercoverage would rise to 100 percent.

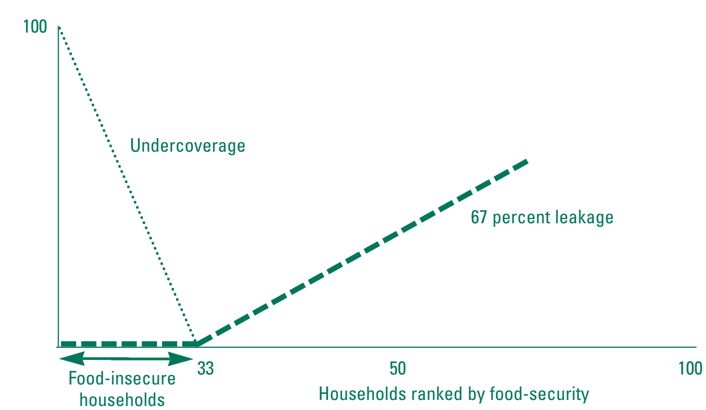
Table 6.4 Errors of inclusion and exclusion under perfect targeting

Participation	Food-insecure	Food-secure	Total
Participate in intervention	Success 33	Inclusion error 0	33
Do not participate	Exclusion error 0	Success 67	67
Total	33	67	100

Source: Compiled by author.

Now consider Figure 6.3. The horizontal axis ranks households on the basis of their degree of food security relative to the total population. Moving from left to right is associated with increasing household food security. Relative to the median household, denoted by "50," all households to the left have lower degrees of food-security and all households to the right have higher degrees of food-security. Recall that 33 percent of the population is food-insecure. The vertical axis measures percentage errors associated with leakage and undercoverage. Now suppose the bottom third of households, as ranked by this indicator, are targeted. For the moment, assume that there is "perfect" information on their food-security status. That is to say, we can identify, without error, who is food-secure and who is food-insecure. In this scenario, undercoverage falls as we move from the first to the 33rd household (as are included more and more food-insecure households), and is zero when 33 or more households participate in the intervention. Leakage is zero when only food-

Figure 6.3 Leakage and undercoverage with perfect targeting



Source: Devised by author.

insecure households are included, but rises as you begin to include households that are not food-insecure.

Now consider the rather extreme example whereby the indicator classifies all food-secure households as insecure and all insecure households as secure. As in the previous scenario, if all households were excluded, leakage would be zero and undercoverage 100 percent. If everyone were included, leakage would be 67 percent and undercoverage zero. If only the bottom third of households were permitted to participate in this intervention, even though none of them are food-insecure, the table of errors of inclusion and exclusion would appear as below (Table 6.5).

Table 6.5 Errors of inclusion and exclusion under "worst case" targeting

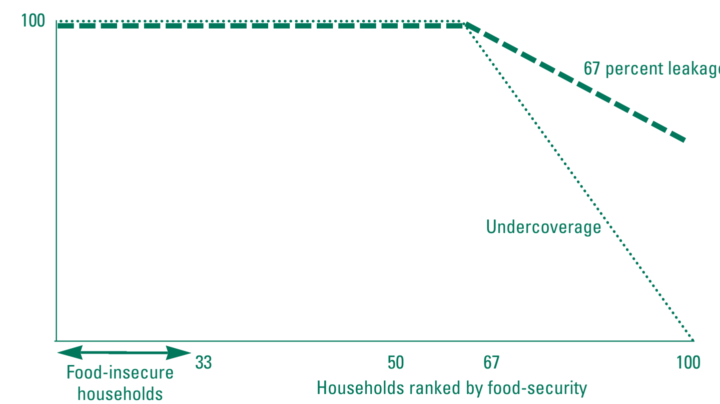
Participation	Food-insecure	Food-secure	Total
Participate in intervention	Success 0	Inclusion error 33	33
Do not participate	Exclusion error 33	Success 34	67
Total	33	67	100

Source: Compiled by author.

Undercoverage is 100 percent of food-insecure households. Undercoverage rates only begin to fall when you double the number of participants. By definition, leakage is zero where there are no participants. It jumps to 100 percent if only one (food-secure) household participates in the intervention and remains at that rate until the 68th household is enrolled in the intervention. This is shown in Figure 6.4.

The purpose of these examples is to illustrate a simple point: Bad targeting has the potential to produce worse outcomes than no targeting at all. It also tempts the project designer to think of

Figure 6.4 Leakage and undercoverage under "worst case" targeting



Source: Devised by author.

targeting issues in terms of the trade-offs between leakage and undercoverage. Consider a more realistic case in which an imperfect indicator of household food security is available. Reducing the number of beneficiaries will reduce leakage, but at the cost of increased undercoverage. How should one assess the trade-off between these? One might be tempted to assume that the objective should be to minimize the sum of leakage and undercoverage, but this is incorrect. Such an objective implicitly measures the success of targeting not in terms of its impact on poverty, but instead on the identity of the recipients. In the literature on the targeting of social programs, where the greatest priority is to improve the welfare of the poor, reducing undercoverage is more important than minimizing leakage. If the priority is to conserve limited budget funds, measures to reduce leakage are given greater weight.

For project staff, the objective of targeting should be seen as the maximization of the reduction in food insecurity, given a fixed budget constraint. The example below develops this further.

An Extended Example

In the Zone Lacustre region of Mali, one of the poorest regions of the world, a number of donors are active in providing resources to improve irrigation facilities so as to better capture water emanating from the annual flooding of the Niger River. By doing so, these projects seek to increase yields of sorghum and rice and to stabilize yields from year to year.

A small field study surveyed 275 households in the Zone Lacustre. These households reside in 10 villages that are grouped geographically into three areas. Data collected at the height of the hungry season indicated that food insecurity is a significant problem. Across all households, caloric availability per person per day was 2,100. Approximately 69 percent of individuals were not obtaining their minimum calorie requirements of 2,200 calories per day.

Would the food-security impact of the ongoing projects in this region be improved by targeting areas, villages, or even households? Answering this question requires an assumption regarding the nature of the intervention. Here, it is assumed that the resources of the project are sufficient to increase the availability of sorghum by 23 kilograms per year for every person in the sample. After taking into account processing, this is equivalent to increasing caloric availability by 100 calories per day. It is further assumed that if this intervention were made available to only half the households in this sample, (approximately) twice the amount of sorghum would be available to each person. That is, it is assumed that no costs are incurred in identifying recipient households. Nor does the cost of the intervention vary with the number of participants. These assumptions are made in order to keep the budget for the intervention fixed. They are relaxed later in the example.

Suppose that no targeting took place. Every household and every person benefits from the intervention, with the result that daily per capita caloric availability rises by 100 calories. This can be called a universal intervention. Table 6.6 summarizes its impact. By definition, undercoverage is zero (every food-insecure person is a participant) and leakage will be 31 percent (every food-secure person is also a participant). The percentage of food-insecure people—the P0 measure—falls by about 2.5 percent. The measure of the severity of food insecurity falls by 15 percent, from 0.157 to 0.132. An alternative to this base case is to assume that households are randomly selected to participate in this intervention. Assuming that one out of every four households is selected makes it possible to increase the impact of the intervention on each individual to 320 calories per day. This is called the random intervention. Leakage is exactly the same as under the universal intervention. Since households are selected randomly, one would expect that 31 percent of the participants would be food-secure; this is precisely what is observed. Undercoverage is 69 percent. Recall that it measures the proportion of individuals who the intervention fails to cover. Given that only 25 percent of households are participants, it is not surprising that undercoverage is much higher under a random intervention than under the universal intervention. But note that the impact on the percentage and severity of food insecurity is virtually identical under either. The intuition for this can be found by going back to Figure 6.1. Although everyone participates under the universal intervention, the benefits to participation are relatively small. Consequently, only those whose existing caloric availability is close to requirements are lifted out of food insecurity. The random intervention, although affecting only a smaller number of individuals, has a larger impact and therefore can pull individuals with lower existing caloric availability out of food insecurity.

Table 6.6 The impact of alternative targeting mechanisms on the percentage and severity of food insecurity

		No targeting			
	Preintervention	Equal allocation to all households	Equal allocation to a random-selected 25 percent of all households	Equal allocation to all persons in the most food-insecure area	Equal allocation to all persons in the four most food-insecure villages
Number of households		275	69	74	63
Number of individuals		1,601	500	470	401
Effect of intervention (expressed in terms of increased daily caloric acquisition)		100 calories per person per day	320 calories per person per day	340 calories per person per day	400 calories per person per day
Leakage		31 percent	31 percent	32 percent	8 percent
Undercoverage		0 percent	69 percent	71 percent	67 percent
P0 (Head-count ratio)	0.694	0.677	0.672	0.685	0.654
P1 (Food-insecurity gap)	0.291	0.260	0.262	0.261	0.252
P2 (Severity of food insecurity)	0.157	0.132	0.133	0.132	0.126

Source: Compiled by author from survey data.

Now suppose the intervention is targeted to the area where average daily per-person caloric availability is the lowest. Even though each person is now receiving three and a half times the amount of calories when compared with a universal intervention, it performs no better in terms of reducing the percentage or severity of food insecurity. By contrast, targeting the four most food-insecure villages generates a slightly better impact on the percentage measure. The P2, or severity measure, is significantly improved, falling 20 percent from the pre-intervention case.

These results come about for the following reason. The most food-insecure area has three villages. One is the most food-insecure village in the sample, the other two are about average. As a result, when targeting this sector, much of the intervention is "wasted," in the sense that benefits go to individuals who are not food-insecure. Targeting the poorest four villages does not produce a much greater impact on the percentage measure because so many people are so far below their requirements. However, it is for exactly this reason that the impact on the severity measure is much higher.

Based on this information, should the intervention be made universally available, allocated randomly, or targeted? It is at this point that the costs of targeting, which up to now have been assumed to be zero, become important. Suppose there is a fixed cost associated with providing this intervention to each household. If this were the case, the random intervention is to be preferred over the universal intervention. Either produces the same impact on all measures of food insecurity, but the random allocation will do so at less cost. Note that the random allocation is preferable, even though the rate of undercoverage is much higher. It is not clear whether the random allocation is preferable to one targeted to the most food-insecure area. Both have comparable rates of leakage and undercoverage, both produce equal reductions in the percentage and severity measures, and both reach about the same number of households. Targeting would only be preferable if the costs associated with obtaining the data necessary to target are more than offset by the possibility that providing the intervention in only one area was cheaper than spreading across an equal number of households scattered over the Zone Lacustre. The case for targeting by village is slightly stronger. Although this might be more costly than targeting the poorest area, it requires reaching a smaller number of households and yields a much larger reduction in the severity of food insecurity. If the savings associated with working in only four villages instead of working with slightly more households spread out over the entire zone outweigh the costs of collecting the information necessary to target, then this form of targeting would be preferable on the basis of its improved impact on the severity of food insecurity.

The Practice of Targeting

In this section, it is assumed that there are gains associated with targeting, either in terms of reducing the cost of providing the intervention or of increasing the impact on food security. The next issue is how to target interventions. Broadly speaking, interventions can be administratively targeted or self-targeted.

Administratively Targeted Interventions

Administratively targeted interventions are those in which project staff determine who will participate. Eligibility is based on a set of criteria. Administrative targeting can be further subdivided into categorical or indicator-based targeting and individual assessment (which includes means testing).

Indicator-based targeting begins by recognizing that means testing may be very costly and may not be entirely accurate. It assumes that there is an identifiable characteristic or set of characteristics that are correlated with, say, food insecurity or poverty. Data on these characteristics are assumed to be relatively easy to obtain. An obvious indicator for project staff to use is geography. Geographical targeting works best when food security differs across regions but is similar within regions. The within-region homogeneity aspect is one that sometimes goes unappreciated. If subregions exhibit great variations in their degree of food insecurity, there is a risk of siting an intervention in a relatively well-off area within a larger, poorer one. For this reason, geographic targeting works best when the geographic units are relatively small districts as opposed to provinces, counties as opposed to states, and so on. Geographic targeting is also attractive on the grounds that it is easier and less expensive to administer. Further, as the example in the previous section illustrated, concentrating resources on fewer units (districts, villages, or households) can have a larger impact on food insecurity.

It is also important to be clear on the criteria used to target geographically. Does attempting to reach the poorest of the poor mean (1) siting interventions where the percentage of food-insecure households is highest (that is, targeting on percentage (P0) measure described in the previous section); (2) siting interventions where there are many food-insecure people (that is, targeting on absolute numbers—sometimes expressed on a density basis); or (3) siting interventions based on the extent of food insecurity among the food-insecure (that is, targeting on severity, the P2 measure described in the previous section)? These different measures will not necessarily yield the same rankings. Again, the Zone Lacustre data can be used to illustrate this. Using the same definition of food security (caloric availability relative to requirements), each village is ranked according to these criteria. The rankings are in descending order of insecurity: a 1 means that the village is the most insecure according to this criterion; a 10 means that it is the most food-secure.

A particularly notable village in Table 6.7 is Gouaty. Based on the criteria developed in the previous section, 95 percent of individuals living in this village are food-insecure, hence its ranking of "1" indicates that it is the most food-insecure village. However, because Gouaty has a very small population, the absolute number of food-insecure people is small compared with other villages in this region. Consequently, it receives a higher ranking using an "absolute number" criterion.

Suppose funding was only sufficient to provide the development intervention in three villages. Using percentage of food-insecure as the criterion, these would be Gouaty, Angira, and Hamakoira. Using absolute numbers, they would be N'goro, Ouaki, and Tomba. Using severity, they would be Angira, Tomba, and Hamakoira. Note that no two rankings produce an identical list of villages.

Table 6.7 Ranking 10 Zone Lacustre villages by percentage of, absolute numbers of, and severity of food insecurity

	Food-insecurity criteria		
Village	Percentage	Absolute numbers	Severity
Aldianabangou	9=	6	5=
Tomba	6	3	2=
Hamakoira	2=	5	2=
Mangourou	8	7	10
Gouaty	1	8=	4
N'goro	5	1	8
Tomi	4	8=	9
Goundam Touskel	9=	10	5=
Ouaki	7	2	5=
Angira	2=	4	1

Source: Compiled by the author from survey data.
Note: 1 denotes the most food-insecure village and 10 the least food-insecure village accoring to this measure; = denotes a tied ranking between two villages.

In the example considered above, villages are ranked on the basis of a single criterion. In practice, project staff may have access to multiple indicators on poverty and household food security. In these circumstances, it is possible to develop a "targeting algorithm"—a statistical method that assigns weights to the relative importance of each indicator. Baker and Grosh (1994) show how this can be done with household data. Morris, Hoddinott, and Coulibaly (2000) illustrate this method using community (village) information.

Targeting at the household level can be done on the basis of indicators (using the algorithm method described above) or means testing. Under means testing, the project obtains information on every potential participant. Based on this information and the criteria for participation, a person is either selected or not selected. A range of methods is available to do this. These are described in Table 6.8.

There are several common problems that affect virtually all methods of targeting of individuals or households. First, indicators of well-being will move over time in response to both transitory and secular, or permanent, phenomena. Targeting on the basis of information collected at a single point in time may include households that are no longer food-insecure and miss households that have fallen into food insecurity. Second, the marginality of poor and food-insecure households manifests itself in many ways, including geographic inaccessibility. Household-level targeting, therefore, will require an aggressive effort to seek out the poorest members of any community.

Table 6.8 Household-targeting mechanisms

Mechanism			
Description	**Advantages**	**Disadvantages**	**Administrative requirements**
Community-based identification Community identifies food-insecure households, for example by using group informant rating (see Chapter 4).	• Simple • Inexpensive	• Communities have incentive to overstate degree of food insecurity. • Therefore triangulation is needed. • Need relatively skilled staff to undertake PRA activity • Since community rankings are relative to community measures, may lead to inconsistencies in terms of access to interventions across communities	• Staff to conduct PRA activities • Record keeping
Household self-reported status Household reports level of food security, for example, by reporting changes in composition, frequency, and size of meals relative to norm.	• Simple • Inexpensive	• Inaccurate • Households have enormous incentive to lie, especially when no triangulating information is collected	• Staff to conduct interviews • Record keeping
Measured food-security status Household food security is measured, for example, via direct observation, 24-hour or 7-day recall. Additional data are collected to adjust for household size and seasonality and to triangulate measured food security.	• Accurate • Difficult for households to offer deceptive information	• Expensive • Lengthy	• Staff to conduct longer interviews • Staff to enter and analyze data • Detailed data entry and record keeping
Proxy measure of food security A synthetic food-security score is calculated on the basis of a set of easily collected indicators of food security, for example, household size, gender of head, diversity of diet and so on. Score can be calculated on-site or by computer.	• Weighting algorithm ensures uniformity in assessment across communities • Not clear to households how to deceive effectively	• Requires longer interview than self-reported status, but shorter than measured food security • Weighting algorithm is inflexible and may not detect special circumstances, for example, natural disasters • May be seen as arbitrary if communities and households do not understand weighting algorithm	• Staff to conduct interviews • Detailed record keeping • Computerized option requires data entry capacity. Software design can be centralized. • Previous analytical work and periodic updates to establish proxy variables and weights

(continued)

Table 6.8 Household-targeting mechanisms (continued)

Mechanism			
Description	**Advantages**	**Disadvantages**	**Administrative requirements**
Fieldworker assessment of food security Subjectively assess same information collected under proxy measurement.	• Can detect special circumstances	• Uniformity and consistency across and within communities hard to maintain • May be perceived as open to favoritism, bribery	• Staff conduct interviews • Record keeping
Nutritional status Anthropometry (see Chapter 2)	• Objective, verifiable, accurate indication of need	• Nutritional status is an outcome of several factors of which food security is only one therefore targeting interventions based on this measure may not be appropriate. • Unclear how households without small children will be assessed • Requires specialist staff to make measurements	• Specialized staff • Computerized analysis of data

Source: Adapted from Grosh 1994.
Note: PRA stands for participatory rural appraisal.

A further problem is the incentive that individuals and communities have to misrepresent themselves in order to increase the likelihood they will be selected for an intervention. This was brought home to the research team undertaking the Zone Lacustre study rather dramatically as part of their participatory rapid appraisal (PRA) work. One component of these exercises was a group-informant rating activity (see Chapter 4). In one village, this activity began with asking the community to provide a local definition of food security. Based on the definition that was eventually agreed upon, the community was then asked to classify households into one of three groups: always food-insecure; sometimes food-insecure; never food-secure. After a brief discussion, the research team was told that all households in this village were always food-insecure. This remark was queried, as a complementary household survey had indicated a certain degree of economic differentiation within this village. Although many households were poor, the household-level data on crop production and livestock indicated that this was not the case for all households. After a protracted discussion, it emerged that these villagers had assumed—despite frequent denials from the research team—that this information would be used as a targeting mechanism, and hence were at pains to disavow any admission of food security in their village.

One can readily imagine such deception—which is carried out for entirely understandable reasons—also occurring when households or individuals are the unit of response. Where targeting is on the basis of characteristics, such as possession of consumer durables, households will have an incentive to hide these. If the indicator is household size, households will have an incentive to amalgamate. One solution to this difficulty is to not reveal the criteria used to select households. But this solution runs counter to the notion of transparency and participation that underlie the interventions of many development agencies. An alternative solution is the use of an objective measure such as child anthropometry. But, as is discussed further in Chapter 2, this is not always appropriate. For example, it excludes households with no children.

Finally, it is worth noting that communities may resist the notion of household or individual targeting. Such an approach might be seen as creating or exacerbating social tensions within villages. The selection of individual households can place local staff in an awkward position, caught between the demands of households to be included and more senior staff that demand adherence to specified targeting criteria.

Self-Targeting

An alternative to administratively targeting interventions is self-targeting. Under self-targeting, the intervention is, in principle, available to anyone who wishes to take part. However, it is designed in such a way that it is only attractive to poorer households. The classic example is that of public works programs that pay a subsistence wage. In the context of most development interventions, examples of self-targeting could include tying the intervention to some time commitment on the part of households. For example, only households who provide labor to a public works program, or who attend meetings with agricultural extension workers or with local health staff, would be eligible to receive a package of interventions such as credit and seeds. The assumption here is that the time costs—in terms of foregone earnings—would be such that only poorer households would want to participate. Alternatively, the intervention could be designed in such a way as to focus on crops or livestock that are especially important for poor people. Research on improved varieties of cassava is an excellent example, since it is a crop consumed primarily by the poor. Self-targeting can also be seen as a means of reaching particular household members such as women. For example, the intervention could be focused on crops grown only by women (though careful design must be undertaken to ensure that these are not then appropriated by men).

Self-targeting requires particularly careful project design. Poorly self-targeted interventions may be attractive to no one, not even the poor. This could occur, for example, where there is no demand for the intervention or where the costs of participation outweigh the benefits. Information on the design of these interventions can be obtained through participatory appraisal techniques, such as the conceptual mapping of threats to food security and SWOT (strengths, weaknesses, opportunities, and threats) analysis, described in Chapter 4.

7. Designing Methods for Monitoring and Evaluating Food Security and Nutrition Interventions

Calogero Carletto and Saul S. Morris

Introduction

In recent years, many development agencies have made intensive efforts to improve their efficiency and increase their impact on rural poverty. At the heart of this new strategic management process is the measurement of performance. With household food security (HFS) and nutritional security now clearly identified as desired outcomes of many development projects, there is a need to assess the performance of investment projects in terms of their impact on household food security and the nutritional status of targeted groups.

When the target populations of development agencies are highly risk-prone, they require rigorous formulation and monitoring. Poorly thought-out evaluations may inadvertently act as an incentive to target better-off groups, who offer higher returns and promise faster disbursement of project resources. In addition, there is a clear danger of placing a higher priority on more easily measurable outcomes or indicators, which fail to provide the information necessary to address broader objectives or to enhance the effectiveness of rural development projects for "the poorest of the poor." Proper evaluations should also reflect an increased awareness that less tangible objectives—such as the formation of social capital, for example—are worthy. Less tangible and broader development objectives do not, however, justify less rigorous evaluation methods. On the contrary, they call for subtler and more sensitive methodologies and indicators.

This chapter emphasizes the design of quantitative impact evaluation exercises for HFS and nutrition. It provides development practitioners with the basic principles on why, when, and how to choose and implement a particular evaluation system. Two of the key features of a good impact evaluation study are the availability of accurate baseline information and a properly thought-out control group, which allows before-after and with-without comparisons. The importance of a joint temporal and cross-sectional comparison of the beneficiary group against a counterfactual is crucial to simultaneously control for the effects of all sorts of external factors likely to contaminate the impact evaluation results. The involvement of the evaluation team in the earliest stages of project design is the most suitable way to ensure a proper and accurate evaluation without having to rely on more complicated statistical techniques, as well as permit a sound learning process to ensue from the evaluation exercise. However, if the conditions dictate, statistical techniques can still provide the evaluation team with effective tools for a well-founded impact evaluation.

Drawing on seminal work recently completed by the UNICEF Evaluation Office, the following sections provide the reader with the conceptual underpinnings for the choice of a particular design suited to the type and level of accuracy of the information required by the different intended end-users. The second part of the chapter, we report on two of the evaluation methodologies used in the field in the course of projects focused on strengthening HFS and nutritional aspects of poverty alleviation projects.

What Kinds of Information Should Be Sought?

A comprehensive evaluation exercise should closely follow the chronological and logical progression of a project cycle. It comprises four sequential steps: an assessment of first the provision, then the utilization, coverage, and impact of new services (Habicht, Victora, and Vaughan 1997). Provision means availability of new services, such as credit lines with commercial banks. Utilization implies the measurement of the rate of use of these services, such as disbursement of loans to smallholder farmers. The issue of coverage asks whether the target population is being reached—for instance, what proportion of poor smallholder farmers has been able to take out a new loan?

The provision of a service by a project, if extended to and properly utilized by a sufficiently large number of beneficiaries, should have an impact on certain variables of interest among the beneficiary population. A number of relations and assumptions link the provision of the service to its impact. A thorough understanding of the existence and strength of these linkages will have a major effect on the form of interventions proposed by the project and, ultimately, on the design of the evaluation system.

The first objective of an evaluation exercise is usually to assess service provision. Once this is done, it may be important to evaluate the level of utilization of such services by the intended beneficiaries and their coverage (take-up) by the project's target groups. It is only when the correct service is provided in a timely manner and properly utilized by a sufficiently large number of beneficiaries that one can plausibly expect an impact on the indicator of interest. Only in these cases is an impact evaluation required or justified.

Project evaluation is a gradual process leading to impact evaluation when the information is required and the conditions call for it. For example, if—based on a preliminary evaluation of the project implementation—one is able to assert with a certain degree of confidence that the provision of the services provided by the project was largely inadequate, or that the level of utilization of the service by the targeted beneficiaries was extremely low, then the situation may not justify pursuing the evaluation further to measure impact. Even in situations in which the project has reached a large group of beneficiaries, and the service has been widely utilized, an impact evaluation exercise may be fruitless if the project has been short-lived or the nature of the intervention is such as to make it unreasonable to expect results within the elapsed time period.

On the other hand, limiting an evaluation to an assessment of service provision, utilization, or take-up based on shaky assumptions about the relationships between project interventions and end-results can be equally improper and misleading. For instance, stopping short of measuring the impact of a small-scale irrigation project on the food security of the households adopting the technology, based on the simplistic assumption that improved irrigation must have had an effect on household agricultural output and access to food, is likely to be inappropriate.

While there undoubtedly are cases in which it is possible to assume that the next link is automatic (that is, if there is provision of services, there will be utilization; if there is utilization, there will be coverage; if there is coverage, there will be impact), this "blind faith" should not be allowed to become routine. Especially for the purpose of impact evaluation, it should only be exercised if the nature of the association between the process and impact indicator is well proven. Most importantly, when a weak point is discovered in the chain—for example, provision did not result in utilization, or coverage did not lead to impact—then the evaluation should include a review of the institutional design of the intervention and the underlying model of the relationship between the intervention and expected impacts.

How Accurate Should the Evaluation Be?

In addition to deciding on what to measure, another issue in the design of the evaluation system is to determine the level of specificity of the information sought, as required by the different stakeholders. Again borrowing terminology from UNICEF (Habicht, Victora, and Vaughan 1997), one can identify three different types of statements reflecting different degrees of confidence end-users may require from the evaluation results: adequacy, plausibility, and probability.

An adequacy assessment simply determines whether some outcome actually occurred as expected; for example, did food security/nutritional status improve? This type of assessment may be particularly relevant when evaluating process indicators such as the provision, utilization, or coverage of a particular project activity, such as the distribution of improved seed varieties. It is less useful for impact evaluation, since it is unable to isolate the effects of the project from those of other concurrent processes, such as whether an observed improvement in yields was due to the provision of improved seed varieties by the project, or instead could be partly or completely attributed to unusually good weather in the area of project intervention. Adequacy assessments are attractive since they do not require working with a control group—for example, farmers who did not receive the improved seed. This greatly reduces the complexity of the data collection activities, but the limitations are obvious.

In contrast to adequacy assessments, plausibility assessments permit determination of whether a given change can actually be attributed to the project, by isolating its effect from all other confounding factors. In the above example, did the improved seed program affect household incomes? By disentangling the project effects from other confounding factors, one can state that the project appears to have had an effect above and beyond the impact of nonproject influences.

The need to control for this confounding arises from the fact that over the project life cycle, it is likely that external factors may contribute, positively or negatively, to changes in outcomes measured among project participants. For example, an observed improvement in child nutritional status over the course of the project could be partly due to an inflow of humanitarian food aid increasing food availability in the area. Similarly, in the context of generalized deterioration, any measurement of project impact would tend to underestimate the true effects, since the project activities may have served as a safety net against concurrent adversity, such as a drought or a drop in food aid. For this reason, the mere comparison of indicators before and after the project is likely to result in misleading results, since it is based on the faulty assumption that the two time periods exhibited similar circumstances except for the presence of the project.

An indispensable feature of plausibility assessments is the use of a control group. Ideally, the control group will exhibit identical characteristics (on aggregate) to the beneficiary group, except for project participation. In reality, this is often not the case, since project participants are rarely chosen at random. There is, therefore, ample potential for project beneficiaries to exhibit characteristics that are systematically different from the control group's. It follows that, in addition to controlling for the external confounding factors potentially affecting the observed trends, it is imperative for a proper impact assessment to also control for this between-group heterogeneity. The potential bias likely to arise from a nonrandom participation scheme is generally referred to as selection bias (or, when project participation is a choice variable, self-selection bias).

Based on the objectives of the evaluation exercise and the constraints dictated by the specific conditions, one can select an internal or an external control group. An internal group is formed by elements in the same area of project intervention who could have joined the program but elected, or were constrained, not to do so. Alternatively, an external control group generally includes those units located in an area not served by the project for whom the option to join was never available. The control for location-specific confounding will be required in this latter case unless one can make a strong case for the assumption of homogeneity.

Finally, probability evaluations can ensure that there is a small, known probability that the differences between project and control areas were due to confounding, to a systematic bias, or to chance. The basis for such a statement is random allocation to project intervention or control status, allowing one to determine with a given probability that the average features of the intervention and control groups are identical. The principle of randomization may appear daunting, but in most circumstances, it is relatively straightforward. In addition to randomization, a probability statement will require adequate statistical power; without this, a probability statement becomes simply a plausibility statement.

A rigorous plausibility or probability evaluation will usually be based on a longitudinal-control design, allowing for both before-after and with-without comparisons. The basis for a longitudinal-control study is the access to baseline information compatible with the objective of the evaluation and the availability of a properly selected control group. Both basic requirements for either a probability or a plausibility inference (for example, baseline information and a control group) call for the involvement of the evaluation team at the onset of the project activities to influence the project development process and ensure an adequate earmarking of project resources to the evaluation activities.

In conclusion, the appropriateness of adequacy, plausibility, or probability evaluation depend on a number of factors such as (1) the objectives of the project, (2) the technical skills of its implementors, (3) the identity and technical sophistication of the end-users of the evaluation results, and (4) resource and time availability. Generally, it may be desirable that new project approaches be rigorously evaluated in at least one location using probability evaluations. In more routine situations, the use of plausibility evaluations may be more cost-effective and sufficient to provide decisionmakers with broad policy options based on a wide spectrum of experiences. Adequacy assessments should be avoided.

What Indicators Should Be Used in the Evaluation?

The choice of indicators to be used in assessing project impact will depend on the stated objectives of the project and on the use to which the evaluation is to be put. This will require identifying the end-users of the information. Also, simplicity and replicability of the indicator may be important. Whenever feasible, the inclusion of both process and impact indicators in the evaluation system will allow different stakeholders to assess a project-induced change in its many dimensions. The question of selection of outcome indicators for nutrition and HFS is extensively discussed in Chapters 2 and 3. Any indicator that is chosen should reflect the true and broad objectives of the project being evaluated.

Avoiding Perverse Incentives in Evaluation of Development Projects

A poorly thought-out monitoring and/or evaluation scheme is likely to create perverse incentives for project managers and implementors alike. The choice of a particular design or set of indicators will almost certainly affect the way that project activities are selected and implemented. A well thought-out monitoring and evaluation scheme is intended to feed into the different stages of project development and contribute to the correct identification of instruments and methods. For instance, it is clear that putting an emphasis on targeting the poorest may affect traditional performance indicators, and as such the results of the project should not be belittled. To take another case, assessing the performance of a credit system solely on types of financial indicators (such as disbursement or repayment rates) is likely to create an incentive for the project management to service primarily better-off beneficiaries, who are the ones most likely to repay. The assessment of the credit program based on the disaggregation of loan disbursement rates by socioeconomic groups of recipients could be a way to partly circumvent this problem. Alternatively, different weights could be assigned to the different groups, creating a bias in favor of more marginal elements. Keeping the evaluation simple by providing more accessible and timely information will allow better monitoring of project activities by project stakeholders. The call for simplicity and rigor of the evaluation system are not conflicting concepts and should be pursued in parallel.

CASE STUDIES

In this section, we provide two practical examples of the impact evaluation methodologies, using case studies from Honduras and Malawi. In Honduras, the low coverage of the program, the existence of a highly comparable, accessible yet geographically isolated control group, and the phased nature of the intervention facilitated the design of a robust "quasi-experimental" evaluation (see Valadez and Bamberger 1994). In Malawi, the lack of baseline information—combined with the nonrandom nature of project intervention—dictated a different approach. Statistical manipulation and data collection techniques based on recall methods were used to control for the potential biases and other confounding. Both evaluation methods relied on a control group and were an attempt to make strong plausibility statements about the projects' impacts on HFS and nutrition. In view of the difficulties involved and the expertise required, the statistical approach used in the Malawi case study must generally be seen as a reserve option for development projects. Early involvement of the evaluation team and the commitment to provide solid technical and financial backing to the evaluation system can help overcome the implementation constraints of the statistical approach.

Honduras

This section describes the evaluation of the interim impact on HFS and nutrition of the Rural Development Plan of the Western Region (PLANDERO) in Honduras, close to the borders with Guatemala and El Salvador. PLANDERO is a rural development project that aimed to increase incomes of the rural poor through transfer of improved technologies, farmer training and credit, and financial services. The evaluation aimed to provide a strong plausibility statement of the degree to which observed changes in both "service" coverage and final outcome indicators could be attributed to project activities. This approach was chosen once it became apparent that random allocation of project interventions between beneficiary and control communities would not be politically feasible in this setting.

The evaluation design took advantage of two fortuitous features of the PLANDERO project: First, the incorporation of beneficiaries was to be phased over several years, and second, the overall coverage of the project did not (and was never intended to) exceed 8 percent (5,000/60,000) of the target group of poor rural households in the area.

The phased incorporation of project beneficiaries meant that even though the evaluation started one year into the execution of the project, it was possible to identify a sample of households (clustered in producers' associations) who were about to receive credit and technical support for the first time. This permitted the collection of "baseline" information, against which the impact of subsequent PLANDERO-related activities could be assessed. Prospective monitoring of this group alone would have been sufficient to identify changes of the type:

$$\Delta_I = I_1 - I_0 , \qquad (7.1)$$

where I_0 is the average HFS or nutrition status of beneficiary households just prior to receiving services or benefits for the first time (referred to as time–0); I_1 is the status of beneficiary households after the introduction of the intervention (time–1); and Δ_I is the observed change, the difference in average HFS/nutrition status between time–0 and time–1. For example, I_0 might be the average dietary energy intake of beneficiary households on incorporation into the project, I_1 might be their average energy intake a year later, and Δ_I would be the difference between the two—a positive quantity where the situation improved, negative where the situation deteriorated, and equal to zero where the situation remained stable. It is, of course, understood that some of the changes identified in the beneficiary group will have been the direct result of project activities, while others are partially due to project activities and partially the result of changes in external factors, and still others are entirely due to changes in the external environment.

Another advantage of the phased incorporation of project beneficiaries in western Honduras was the existence of a pool of communities that had already been earmarked for inclusion in PLANDERO at a later date. These communities generally had functioning farmers' associations, and many of them were already well known to the technical assistance companies that delivered services under contract from PLANDERO. From this pool of "reserve" communities, it was possible to identify a second sample of households (also clustered in producers' associations, for the most part) that were similar to the intervention group I, could also be observed from time–0 to time–1, but would not over this period benefit from the technical assistance or credit available through PLANDERO. This group of communities is referred to as the control communities, *C.* Monitoring of this group permitted estimation of the parameters:

$$\Delta_C = C_1 - C_0 , \qquad (7.2)$$

where C_0 is the average HFS or nutrition status of control community households at time–0; C_1 is the status of control community households at time–1; and Δ_C , the difference between these two quantities, may be interpreted as a measure of the changes due to factors external to PLANDERO affecting outcomes observed from time–0 to time–1.

Valadez and Bamberger (1994, 235–237) have shown that if it can additionally be assumed that (1) the control and intervention communities were similar at time–0, (2) the external factors affecting the control communities and the intervention communities are the same, and (3) the effects of the program are strictly limited to the intervention communities, *I*, then the impact of the project can be estimated as:

$$\Delta_I - \Delta_C = I_1 - I_0 - C_1 - C_0 . \qquad (7.3)$$

In the Honduras case study, great care was taken at the design stage to ensure that the control communities were similar to the intervention communities at time–0: In fact, they were matched one-to-one on the basis of geographical area, altitude, and production system. The intervention and control communities were similar, though not identical, at time–0, as shown below.

	Intervention communities n = 193	Control communities n = 189
Household size (mean/std. deviation)	6.6 (2.7)	6.7 (2.7)
Asset score (mean/std. deviation)	2.0 (0.7)	2.0 (0.7)
Landownership, hectares (mean/std. deviation)	3.0 (4.5)	2.8 (3.5)
Total cultivated area, hectares (mean/std. deviation)	2.1 (2.1)	2.0 (1.9)

Because of the geographical proximity of the control and intervention communities, most of the external factors affecting their food security status would have been uniform. For example, the unusual weather patterns (attributable to the El Niño phenomenon) observed between time–0 and time–1 were common to the entire study area. Similarly, the government's decision to import large quantities of maize in mid-1997 resulted in large drops in the maize price between the 1996–97 harvest and the subsequent one in 1997–98 (median maize prices fell in the intervention communities from L.150/quintal to L.115/quintal, and in the control communities from L.150 to L.110). This design is, however, vulnerable to idiosyncratic changes affecting single communities. It can only be hoped that, by including a large number of different communities in the control sample, the net effect of the sum of all such idiosyncrasies will be zero.

The fact that controls were selected from different communities from the intervention sites had the advantage that there was little contamination of the control communities by project activities (at least during the one-year period observed), which would have the undesirable effect of diluting the apparent project impact. Use of radio programs meant that a certain amount of contamination did occur: For example, even in the control communities, 6 percent of respondents heard about Integrated Pest Control through PLANDERO, and 13 percent heard about Rural Savings Associations from this source. Nonetheless, the ability of farmers in the control communities to transform this knowledge into practice without the support of the project remained limited.

The importance of including a control group can be seen by considering the case of dietary energy intake. From time–0 to time–1, energy intake in the intervention group increased by just 1.6 percent, a negligible change. However, over the same period, energy intake in the control communities fell by 6.0 percent, suggesting that members of the PLANDERO group may in fact have been protected from a small deterioration in energy intake affecting other households in the region. Although the net effect of the project on energy intake from time–0 to time–1 (+7.7 percent) did not attain statistical significance, it should certainly be borne in mind that were this trend to continue for the remainder of the project life cycle, the end result could be significant, both statistically and substantively.

In order to try and compensate for the short period of observation between time–0 and time–1, the Honduras case study supplemented the basic longitudinal-control design with an additional group of households (denoted I–) belonging to producers' associations that had already been receiving technical assistance and credit from PLANDERO for a whole year before the beginning of the evaluation period. Once again, these communities were individually matched to control communities and new-intervention communities on the basis of geographic location, altitude, and production system. The main purpose of including this group of households in the evaluation was to determine whether some of the differences identified between the new-intervention group and the control group could be expected to be maintained or even increased over time. For example, it was found that adoption of the agronomic practices promoted by PLANDERO increased with duration in the program (all contrasts statistically significant at the P<0.05 level), as shown below.

Interventions	Households reporting adoption of recommended practice		
	Control communities	1 year with PLANDERO	2 years with PLANDERO
Rural savings associations	39	69	76
Vaccination of fowl and swine	23	30	45
Correct density of seeds	78	87	89
Hedges	57	71	75
Integrated pest control	35	44	49
Organic fertilizers	39	51	53

Source: Compiled by author from survey data.

It was also assumed that some of the changes resulting from project activities would take time to manifest themselves, perhaps only becoming evident after a "latent" period of intensive training and opinion forming. Thus, it was noted that between time–0 and time–1, households in the control communities increased their total dietary diversity by approximately one food item, while those in the intervention communities increased their diversity scores by 2.5 items on average, and those households in their second year with PLANDERO increased their scores by fully five items (trend statistically significant at the P<0.05 level).

The Honduras case study was fortunate that the matching procedure employed appeared to result in broadly comparable intervention and control groups at time–0. However, there is always a concern that the intervention and control groups may, in fact, have been on different trajectories prior to time–0, intersecting only temporarily at this time; thus, what appeared to be a project impact at time–1 could, in fact, have been nothing more than the inevitable crossing of preexisting trajectories in the control and intervention communities. This topic is dealt with extensively in Valadez and Bamberger (1994, 245–246), and requires monitoring over a longer period of time than was possible in Honduras to convincingly purge the results of the "burden of history." Worse still, the two groups could be similar on the n variables observed at time–0, but quite different on scores of others that were not or could not be observed. As mentioned previously, one way of avoiding this problem is to assign communities to intervention and control groups at random, a process that involves little extra effort when—as in the Honduras case—there is a large pool of "reserves" that cannot, in any case, all be served in the first phase of project implementation. There are, however, formidable political constraints to random allocation of communities, which may, as was the case in Honduras, prove to be insuperable.

Malawi

The Malawi case study examines the impact of a smallholder food security project in central Malawi. The intervention sought to raise farm incomes and reduce food insecurity and child malnutrition via the creation of "joint liability credit clubs" that facilitated access to credit, inputs and markets for maize and tobacco. The project presented a less-than-optimal scenario from the evaluator's perspective for several reasons:

- There was no adequate baseline study of the project area;
- project beneficiaries could not be considered a random selection of all households in the area;
- significant changes in the economic environment occurred following project inception;
- there were no up-to-date data at the village or section level to allow the identification of comparable external control areas;
- there was no up-to-date sampling frame available for nonbeneficiary households; and
- time and resource constraints made it infeasible to construct a comprehensive sampling frame.

Despite these factors, evaluation was still possible, though the technical requirements (in terms of the use of more advanced statistical techniques and software) were higher as a consequence. In the face of these limitations, a number of techniques were used to ensure a reliable evaluation study, including

- use of recall methods to reconstruct the situation of both beneficiaries and control in the preproject period, thus permitting "before-after" comparisons;
- the application of so-called "two-stage estimation procedures" to control for differences between beneficiaries and nonbeneficiaries, arising from the nonrandom selection of beneficiaries from the population (selection bias);
- choice of an internal control group, thus eliminating the need to obtain information on nonbeneficiary communities; and
- use of EPI cluster-sampling methods to identify a representative control group in the absence of a comprehensive sample frame (see Chapter 5).

One of the objectives of the Malawi case study was to assess the impact of project participation between time–0 (time of project onset) and time–1 (time of the evaluation) on a set of chosen indicators. Random allocation of the project intervention would have supported the hypothesis of homogeneity between participants and nonparticipants at time–0 by ruling out the possibility that a biased (self-) selection process was at play. However, as a result of the nonrandom allocation of project resources, project beneficiaries were likely to have exhibited characteristics at time–0 that were systematically different from the rest of the population. Thus, a straight comparison of beneficiary and control groups at time–1 would almost certainly have been biased and would have led to misleading estimates of project impact. It was therefore imperative to control for the potential selectivity bias in the analysis.

Availability of baseline information describing the two groups in the preproject period would have made the estimation procedure more straightforward and accurate had the proper information been collected at time–0. In the Malawi case study, the lack of baseline information made it necessary to use recall methods at the data collection stage for both the beneficiary and control groups to permit the before-after comparison. The possibility of constructing a "longitudinal" data set from a cross-section at time–1 depends on the length of the recall period (that is, the time elapsed between time–0 and time–1), as well as the particular data collection techniques used, the nature of the variables of interest, and the availability of technical expertise and trained personnel in the field to elicit this type of historical information. Fortunately, the relatively recent implementation of the project reduced the difficulties of using recall methods. The magnitude of the problem would have been much larger if the evaluation team had needed to reconstruct historical data going back several years.[1] The prior identification of household- and location-specific temporal benchmarks facilitated the work of the enumerators in assisting the respondent in the recall process. Even so, the dangers and difficulties of reconstructing historical data should not be underestimated.

In this study, the use of an external control group was not deemed appropriate. This was because of the lack of information at a spatially disaggregated level that would have allowed the identification of a control group outside of the project area with characteristics similar to the beneficiary group, eliminating the need to control in the analysis for location-specific differences between beneficiaries and controls. In addition, because of the low project coverage within the area of intervention, the main disadvantage of using an internal control group (potential project spillover effects to nonparticipants) was considered negligible. Because of the lack of sampling frame and the need to reduce the likelihood of a biased selection of the control group, a variant of the EPI cluster-sample design was used to select the control group (for more details on the approach, please refer to Chapter 5).

To illustrate the approach used in the Malawi case study and highlight the consequences of ignoring selection bias, the following paragraphs guide the reader through a fully worked out example of the estimation of the project impact on a selected indicator of nutritional status, height-for-age Z–score of children 6–60 months of age (for more details on this and other similar indicators, see Chapter 2).

The table below shows a straight comparison of the prevalence of stunting between children of participant households and those from the control group suggests that there was essentially no difference in the prevalence of stunting between project children and control group children. This initial result should in no way be interpreted as evidence of the lack of project impact. Multivariate analysis (statistical modeling) would provide more appropriate evidence by accounting in a simultaneous fashion for more than one determinant of nutritional status.

Prevalence of Stunting	Project	Control	Both
Number of observations	111	153	264
Percent stunting (HAZ<–2)	52	56	54

One such method often (but improperly) used is to estimate a multiple linear regression model using ordinary least square (OLS) estimation methods where we include the height-for-age Z–scores as the response variable and a binary (yes/no) variable reflecting participation in the program as one of the explanatory variables, together with all other variables believed to determine child nutritional status. This type of model can be run easily using virtually any statistical package or spreadsheet application. The table below reports the estimated coefficient and the standard error of the participation variable estimated coefficients (after controlling for several child, mother, and household characteristics) for the Malawi case study.

Determinants of Child height-for-age Z–score	Coefficient	Standard error
Project participation	–.022	.18

Just as in the straight comparison, the result suggests (erroneously) that project participation has no effect on child nutritional performance. Because participation in the project is not random, however, the estimate of this coefficient is inaccurate. For a household to join the project, it must satisfy a set of restrictive eligibility criteria. Yet not all households meeting these criteria would decide to join the program. This selection rule indicates that there are both selection (by the project, based on eligibility criteria) and self-selection (by beneficiaries, who elect to enter the program based on some idiosyncratic selection rule, such as expected returns or lack of alternative credit sources) processes at play that invalidate the OLS results. To account for the nonrandomness of this selection rule, you first estimate the probability for a household to join the program by regressing the participation variable on a number of regressors believed to have affected the selection rule. The Probit results yield an estimated variable, the Inverse Mills Ratio (IMR), which can broadly be interpreted as a variable capturing all those unobservable characteristics potentially having an effect on the final outcome (nutritional status), and which differentiate the two groups beyond the project effect.

In order to run the Probit model, you choose a number of variables likely to be associated to the decision to join the program, but that are uncorrelated to child nutritional status. As previously noted, being determinants of the participation decision, the variables included in this first-stage equation should reflect household circumstances at time–0; that is, before the onset of the project activities. As applied in the example, options for the evaluation team include: (1) using postproject variables that are unlikely to have changed during the course of the project life cycle, such as educational level of the household head; (2) using recall methods to collect information on preproject status that are relatively easy for the respondent to remember, such as household composition, sale/purchase of major assets, or cropping patterns of major crops; or (3) a combination of the two. Caution should be used when selecting variables observed at time–1, because there is always a risk that these may have been affected by the program. For example, while landownership at time–0 may be an appropriate choice of variable to explain program participation, the same variable at time–1 may give rise to biased estimates if participation in the program has affected household land accumulation patterns between time–0 and time–1. Examples of a good variable to use in this first stage would be whether the household head previously knew the extension workers in charge of promoting project membership, or whether any relative or friend had already joined the project. Conversely, examples of less appropriate variables would be women's education or a wealth proxy such as availability of a latrine, since although likely to reflect the household human and capital wealth (possible determinants of participation), they are both also likely to be related to children's nutritional performance. It should be noted that with these methods, failure to identify variables that correctly predict project participation will prevent estimation of project impact. Therefore, choosing these variables requires careful planning before the beginning of data collection activities, and demands familiarity both with the method and with local conditions.

The estimated IMR is then included in a second-stage equation that looks exactly like the first OLS equation, except for the added selectivity variable. This second equation can safely be estimated by OLS. The estimated coefficients of the participation and selectivity (IMR) variables (after controlling for several child, mother, and household characteristics) are reported below.

Determinant of Child height-for-age Z–score	Coefficient	Standard error
Project participation	1.07	.50
Selectivity bias (IMR)	–0.75	.33

A test for presence of bias in the program selection process is a test on the coefficient of the IMR. The negative significant value of the coefficient (–0.75) reflects the existence of a negative selectivity bias against participants and a positive selectivity bias in favor of nonparticipants, indicating that project participants exhibit unobservable characteristics that were not inducive of good nutrition, and that, perhaps because of that, were purposively selected into the program.

Under the assumption that the selection model (the first-stage Probit) is correctly specified, the coefficient on the participation variable (1.08) now reflects the true impact of project participation on nutritional performance. The result is quite striking: The estimated coefficient is quite large in magnitude and strongly statistically significant. Participation in the project appears to be associated with an improvement in the height-for-age Z–score of preschoolers by one whole point (one standard deviation). The interpretation of the results is that the project was successful in targeting the worse-off households (as reflected in the negative coefficient on the selectivity variable) and, within the elapsed project cycle, in raising the nutritional performance of their preschoolers to the level comparable to the one exhibited by the control children.

In summary, this is shown a simple example in which correcting for selectivity bias has important consequences for the results. It is also clear that the methods required for this correction demand considerable technical expertise, and are unlikely to be well suited for routine use in country.

Endnote

1. Even in these cases, however, going back a few additional years in the recall of major events is recommended since they may still have an influence on the household's decision at time–0. In the Malawi case study, although the project has been operational for only two years, the recall period went back up to 7 years for some variables.

References

Abel, N., and M. Stocking. 1979. Rapid aerial survey techniques for rural areas. Paper for the RRA conference, December 4–7, held at International Development Studies, University of Sussex, Brighton, U.K.

Bergeron, G., and S. J. Scherr. 1996. Participatory resource mapping for policy research: An application in the hillsides of Central Honduras. Fragile Lands Program Working Paper. Environment and Production Technology Division. Washington, D.C.: International Food Policy Research Institute.

Bernard, H. R. 1988. *Research Methods in Cultural Anthropology.* Newbury Park, Calif., U.S.A.: SAGE Publications.

Bernard, H. R., P. Kilworth, D. Kronenfeld, and L. Sailer. 1984. The problem of informant accuracy: The validity of retrospective data. *Annual Reviews in Anthropology* 13: 495–517.

Borgatti, S. 1993. Methods manual. Anthropac 4.1. Colombia, S.C., U.S.A.: Analytic Technologies.

Casley, D. J., and K. Kumar. 1988. The collection, analysis, and use of monitoring and evaluation data. Baltimore, Md., U.S.A.: Johns Hopkins University Press.

Casley, D. J., and D. A. Lury. 1987. *Data Collection in Developing Countries.* Oxford, U.K.: Clarendon Press.

Chambers, R. 1994. The origins and practice of participatory rural appraisal. *World Development* 22 (July): 953–969.

Childress, L. M., D. J. Herrmann, and S. Schechter. 1995. The use of heuristics in answering questions about past events. Mimeo.

Christiaensen, L., J. Hoddinott, and G. Bergeron. 2001. Comparing village characteristics derived from rapid appraisals and household surveys: A tale from northern Mali. *Journal of Development Studies.* 37: 1–20.

Christiaensen, L., and E. Tollens. 1995. Food security: From concept to action—A status quaestionis. K. U. Lueven and B.A.D.C. research project in the framework of policy preparing research in the field of cooperation in development, No. 2, Brussels.

Chung, K., L. Haddad, J. Ramakrishna, and F. Riely. 1997. Identifying the food-insecure: The application of mixed method approaches in India. Washington, D.C.: International Food Policy Research Institute.

Clay, E. 1997. Food security: A status review of the literature. Research Report ESCOR No. R5911. London: Overseas Development Administration.

Cochran, W. G. 1997. *Sampling Techniques.* New York: John Wiley.

CTA/ECSA (Technical Centre for Agricultural and Rural Cooperation/East, Central, and Southern Africa Food and Nutrition Cooperation). 1987. Food composition table. The Netherlands: Wageningen Agricultural University, Department of Human Nutrition.

Deaton, A., and M. Grosh. 1998. The consumption module in the living standards measurement survey. In *Designing Household Survey Questionnaires for Developing Countries*, ed. P. Glewwe and M. Grosh. Washington, D.C.: World Bank.

Deaton, A., and C. Paxson. 1998. Economies of scale, household size and the demand for food. *Journal of Political Economy* 106 (October): 897–930.

Deutscher, I. 1973. *What We Say/What We Do: Sentiments and Acts.* Glenview, Ill., U.S.A.: Scott, Foresman and Cie.

Devereux, S., and J. Hoddinott, eds. 1992. *Fieldwork in Developing Countries.* London and Boulder, Colo., U.S.A.: Harvester Wheatsheaf and Lynne Rienner.

Elinson, J. 1963. Methods of sociomedical research. In *Handbook of Medical Sociology,* ed. H. E. Freeman, S. Levine, and L. G. Reeder. Englewood Cliffs, N.J., U.S.A.: Prentice Hall Inc.

Engle, P. L., and J. B. Lumpkin. 1992. How accurate are time-use reports? Effects of cognitive enhancement and cultural differences on recall accuracy. *Applied Cognitive Psychology* 6: 141–159.

Erasmus, C. J. 1952. Changing folk beliefs and the relativity of empirical knowledge. *Southwestern Journal of Anthropology* 8: 411–428.

FAO (Food and Agriculture Organization of the United Nations). 1997. Guide for the conduct of the constraints analysis component. Special Programme for Food Security, Handbook Series, SPFS/DOC/18. Rome.

Frankenberger, T., A. Peña-Montenegro, S. Tilakaratna, N. Velarde, and W. B. Eide. 1993. Rural poverty alleviation and nutrition: IFAD's evolving experiences. Staff Working Paper 14. Rome: International Fund for Agricultural Development.

Freeman, L., A. K. Romney, and S. Freeman. 1987. Cognitive structure and informant accuracy. *American Anthropologist* 89: 310–325.

Gibson, R. S. 1990. *Principles of Nutritional Assessment.* New York: Oxford University Press.

Gill, G. J. 1992. Policy analysis for agricultural resources management in Nepal. A comparison of conventional and participatory approaches. Research Support Series No. 9. HMG Ministry of Agriculture/Winrock International. Policy Analysis in Agriculture and Related Reproduce Management. Kathmandu, Nepal.

Greene, W. H. 1993. *Econometric Analysis* 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Gupta, A. K., and IDS (International Development Studies) Workshop. 1989. Maps drawn by farmers and extensionists. In *Farmer First: Farmer Innovation and Agricultural Research,* ed. R. Chambers, A. Pacey, and L. A. Thrupp. London: Intermediate Technology Publications.

Habicht, J. P., R. Martorell, C. Yarbrough, R. M. Malina, and R. E Klein. 1974. Height and weight standard for preschool children: How relevant are ethnic differences in growth potential? *Lancet* 1: 611–615.

Habicht, J.-P., C. G. Victora, and J. P. Vaughan. 1997. Linking evaluation needs to design choices: A framework developed with reference to health and nutrition. UNICEF Staff Working Papers: Evaluation and Research Series, Number EVL–97–003. New York: United Nations Children's Fund (UNICEF).

Hardin, R. 1982. *Collective Action.* Baltimore, Md., U.S.A.: Johns Hopkins University Press.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.

International Fund for Agricultural Development (IFAD). 1998. Annual report, 1997. Rome.

James, W., and E. Schofield. 1990. *Human Energy Requirements: A Manual for Planners and Nutritionists.* Oxford, U.K.: Oxford University Press.

Kish, L. 1965. *Survey Sampling.* New York: John Wiley.

Kroeger, A. 1983. Health interview surveys in developing countries. A review of methods and results. *International Journal of Epidemiology* 12 (4): 465–481.

Lindhult, M. S., J. Fabos, P. Brown, and N. Price. 1988. Using GIS to assess conflicts between agriculture and development. *Landscape and Urban Planning* 16 (4): 333–343.

Maxwell, D. 1996. Measuring food insecurity: The frequency and severity of "coping strategies." *Food Policy* 21: 291–303.

Maxwell, S., and T. Frankenberger. 1992. Household food security: Concepts, indicators, measurements. International Fund for Agricultural Development (IFAD) and UNICEF, Rome.

McNabb, S. L. 1990. The use of inaccurate data: A methodological critique and applications of Alaska native data. *American Anthropologist* 92: 116–129.

Newbold, P. 1988. *Statistics for Business and Economics,* 2nd edition. Englewood Cliffs, N.J., U.S.A.: Prentice Hall.

Olson, M. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, Mass., U.S.A.: Harvard University Press.

Radimer, K., C. Olson, and C. Campbell. 1990. Development of indicators to assess hunger. *Journal of Nutrition* 120: 1544–1548.

Republic of Honduras, Secretary for Education. 1996. *Sixth Census of the Height of First Grade School Students, 1996.* Tegucigalpa: Republic of Honduras, Secretary for Education.

Republic of Honduras, Secretary for Health. 1996. *National Micronutrient Survey 1996.* Tegucigalpa: Republic of Honduras, Secretary for Health.

Riely, F., and N. Moock. 1995. Inventory of food security impact indicators. In *Food Security Indicators and Framework: A Handbook for Monitoring and Evaluation of Food Aid Programs.* IMPACT, Arlington, Va., U.S.A. Draft.

Riely, F., N. Mock, B. Cogill, L. Bailey, and E. Kenefick. 1995. Food security indicators and framework for use in the monitoring and evaluation of food aid programs. IMPACT: Food security and nutrition monitoring project, Arlington, Va., U.S.A.

Rogers, B. L., A. J. Swindale, and P. Ohri-Vachaspati. 1996. Determinants of Household Food Security in Honduras. A report on the National Household Consumption, Income, Expenditure and Nutrition Survey, 1993–94.

Washington D.C.: United States Agency for International Development (USAID).

Romney, A. K., S. C. Weller, and W. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88: 313–339.

Ross, D. A., and P. Vaughan. 1986. Health interview surveys in developing countries: A methodological review. *Studies in Family Planning* 17: 78–94.

Ryan, G. W., and H. Martinez. 1996. Can we predict what mothers do? Modeling childhood diarrhea in rural Mexico. *Human Organization* 55 (1): 47–57.

Smil, V. 1994. How many people can the earth feed? *Population and Development Review* 20: 255–292.

United Nations Administrative Committee on Coordination Sub-Committee on Nutrition. 1998. Third Report on the World Nutrition Situation. Geneva: United Naitons ACC/SCN.

United States Department of Agriculture (USDA), Agricultural Research Service. 1999. USDA Nutrient Database for Standard Reference, Release 13. Nutrient Data Laboratory Home Page.

Uraivan, Tan-Kim-Young. 1992. Participatory land use planning as a sociological methodology for natural resource management. Reproduce Management and Development Program in Chiang Mai University, Thailand Royal Forest Department, and Ford Foundation.

Valadez, J., and M. Bamberger. 1994. Monitoring and evaluating social programs in developing countries. Washington, D.C.: World Bank.

Van Willigen, J., and B. deWalt. 1985. Training manual in policy ethnography. American Anthropological Association (Special Publication) 19.

World Health Organization (WHO). 1985. Energy and protein requirements. Technical Report Series 724. Geneva.

World Health Organization (WHO) 1995. Expert Committee on Nutrition. Physical status: Uses and interpretation of anthropometry. Geneva: WHO.

World Health Organization (WHO)–United Nations Children's Fund (UNICEF). 1994. Indicators for assessing vitamin A deficiency and their application in monitoring and evaluating intervention programmes. A report of a joint WHO/UNICEF consultation. Geneva: WHO.

CONTRIBUTORS

GILLES BERGERON
Academy for Educational Development
gbergero@aed.org
Formerly Research Fellow, Environment and Production Technology Division, International Food Policy Research Institute

CALOGERO CARLETTO
World Bank
ccarletto@worldbank.org
Formerly Visiting Research Fellow, Food Consumption and Nutrition Division, International Food Policy Research Institute

JOHN HODDINNOT
Dalhousie University, Department of Economics
john.hoddinott@dal.ca
Formerly Research Fellow, Food Consumption and Nutrition Division, International Food Policy Research Institute

SAUL S. MORRIS
London School of Hygiene and Tropical Medicine
saul.morris@lshtm.ac.uk
Formerly Research Fellow, Food Consumption and Nutrition Division, International Food Policy Research Institute

Photo Credits:
Page viii, William Spidle
Page 10, World Bank/Curt Carnemark
Page 30, World Bank/Curt Carnemark
Page 46, Ray Witlin
Page 76, World Bank
Page 88, World Bank/Curt Carnemark
Page 102, World Bank